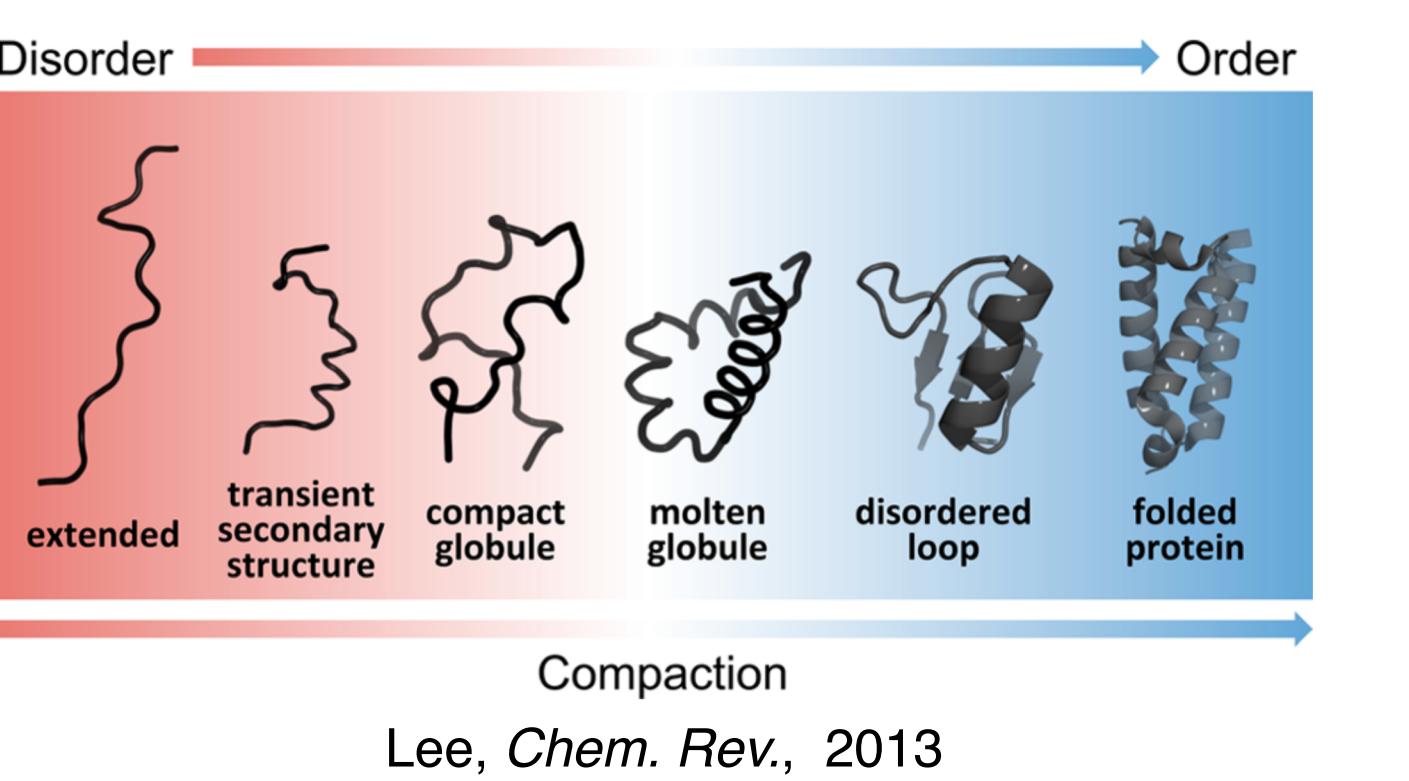
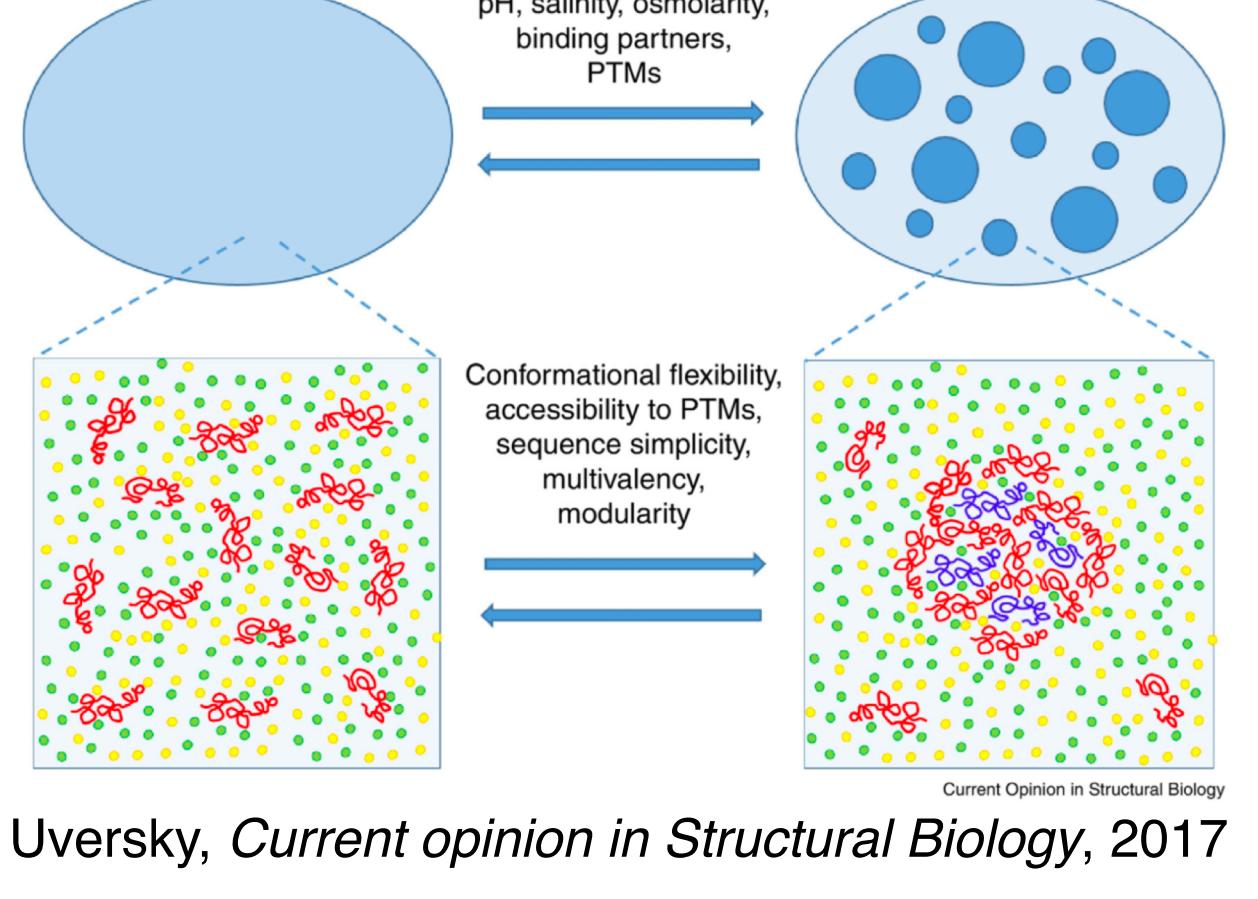


## Question: how are intrinsically disordered proteins linked to Alzheimer's disease?

Intrinsically disordered proteins (IDPs) are functional even though they are unable to fold into a stable conformation. They are important for regulating many cellular processes including transcription, translation, and cellular signaling.

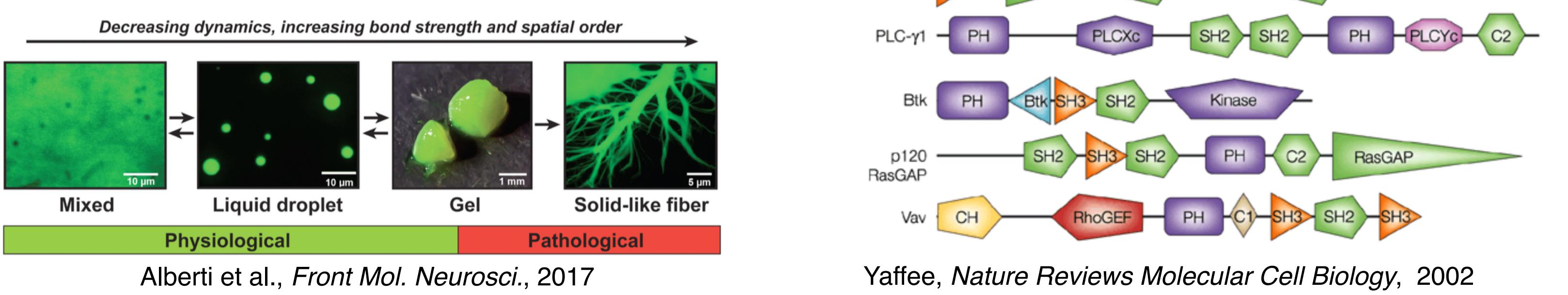


Intrinsically disordered regions (IDRs) (often in tandem with structured regions) give proteins the ability to generate membrane-less organelles to localize biochemical processes in a process referred to as liquid-liquid phase separation.



Uversky, Current opinion in Structural Biology, 2017

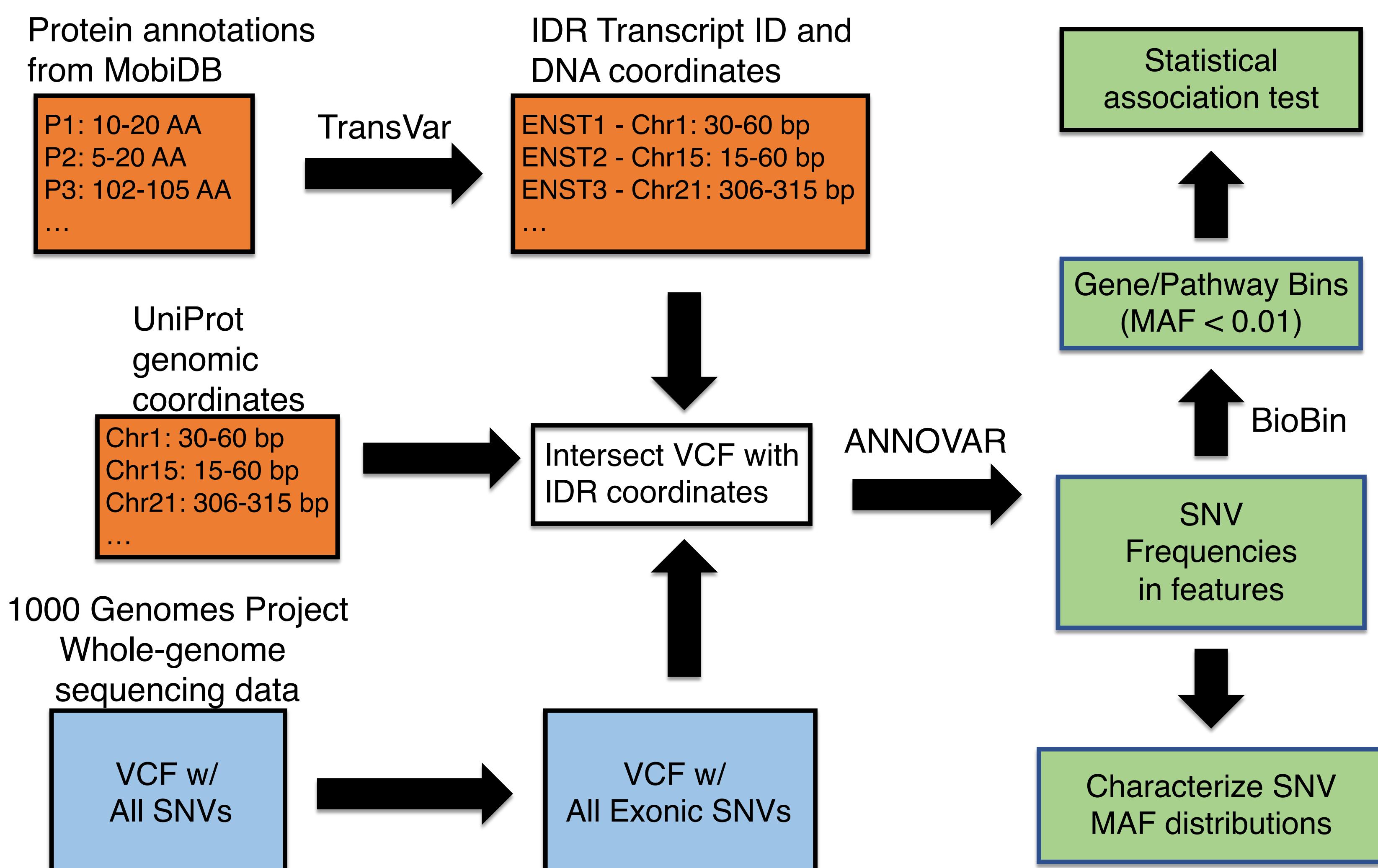
The natural ability of IDRs to mediate interactions with other proteins and generate phase transitions also underlies its connection to protein aggregation in multiple forms of neurodegeneration, including Alzheimer's disease.



Alberti et al., Front Mol. Neurosci., 2017

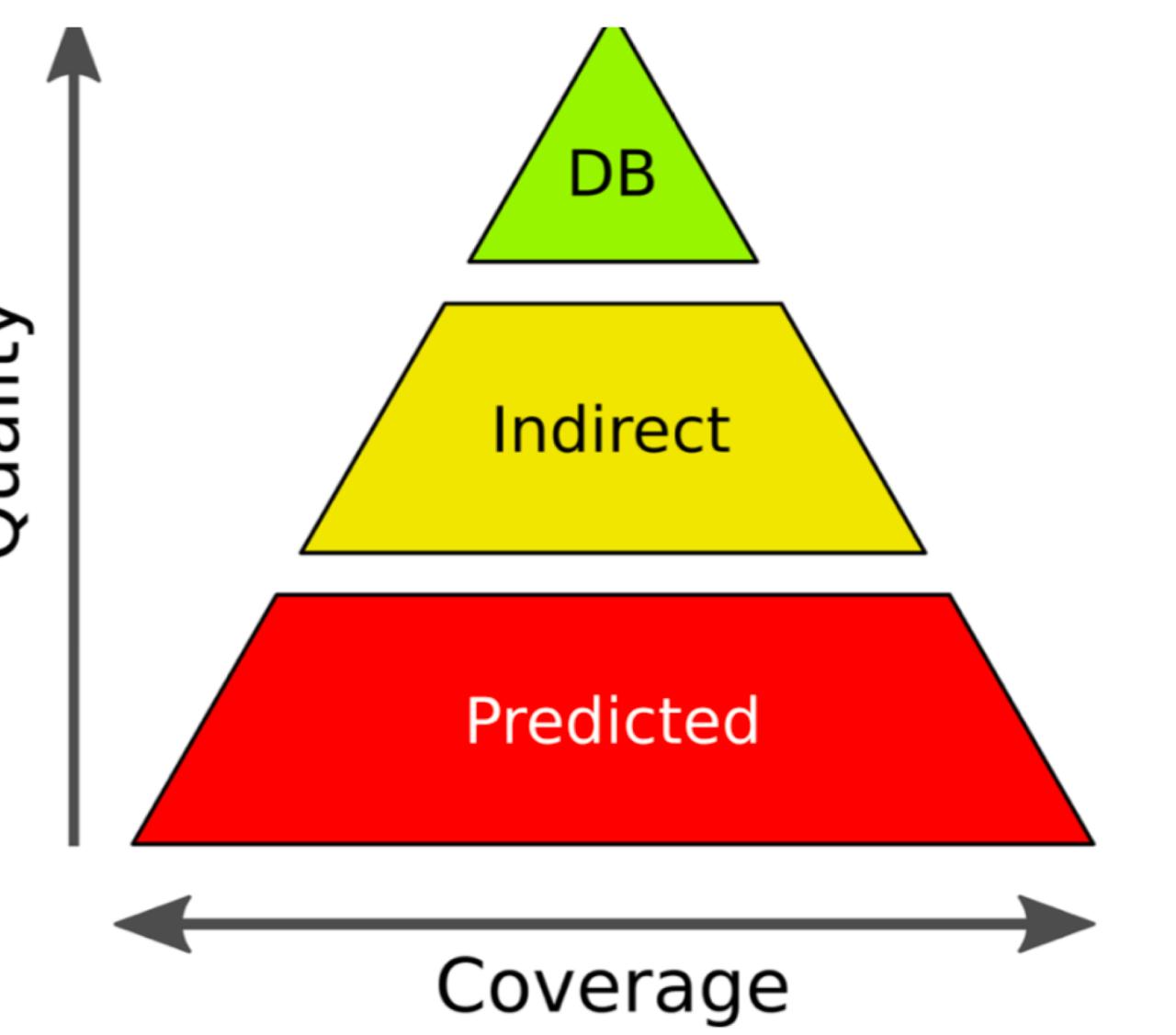
**Hypothesis:** characterizing genetic variation in IDRs and structured domains will lead to the identification of new AD risk loci.

## Linking protein regions to DNA coordinates



## Overview of MobiDB: a comprehensive database of IDR annotations

MobiDB is a database that contains three broad types of IDR information.



Piovesan et al., Nucleic Acids Research, 2018

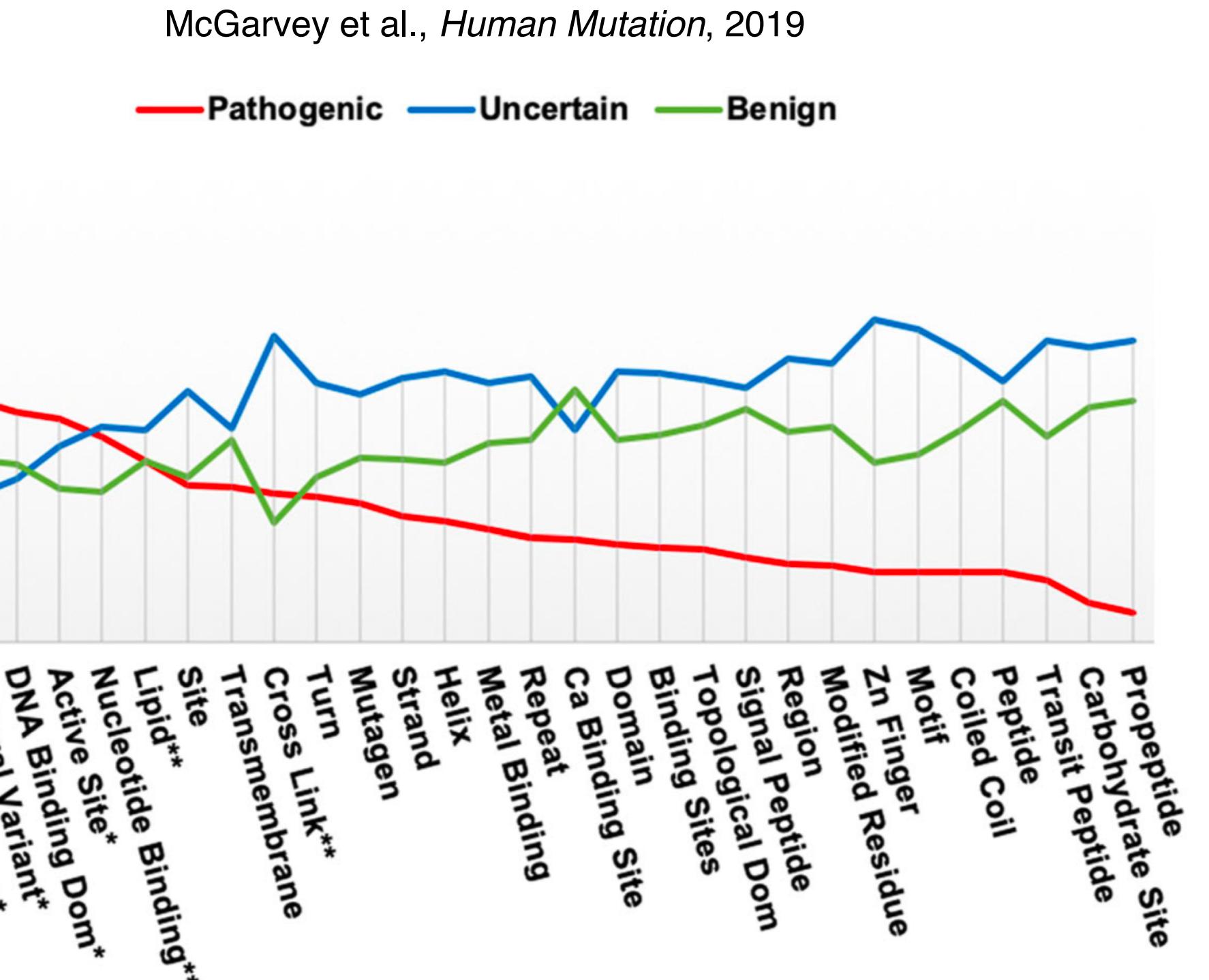
DB: manually curated structural information from databases containing experimental results (e.g. DisProt, FuzDB, and UniProt).

Indirect: information derived from PDB data using missing residues, high temperature residues (X-ray), and mobility (NMR).

Predicted: disorder prediction algorithms that use amino acid sequence information.

UniProt database contains genomics coordinates for structured domains and some IDRs.

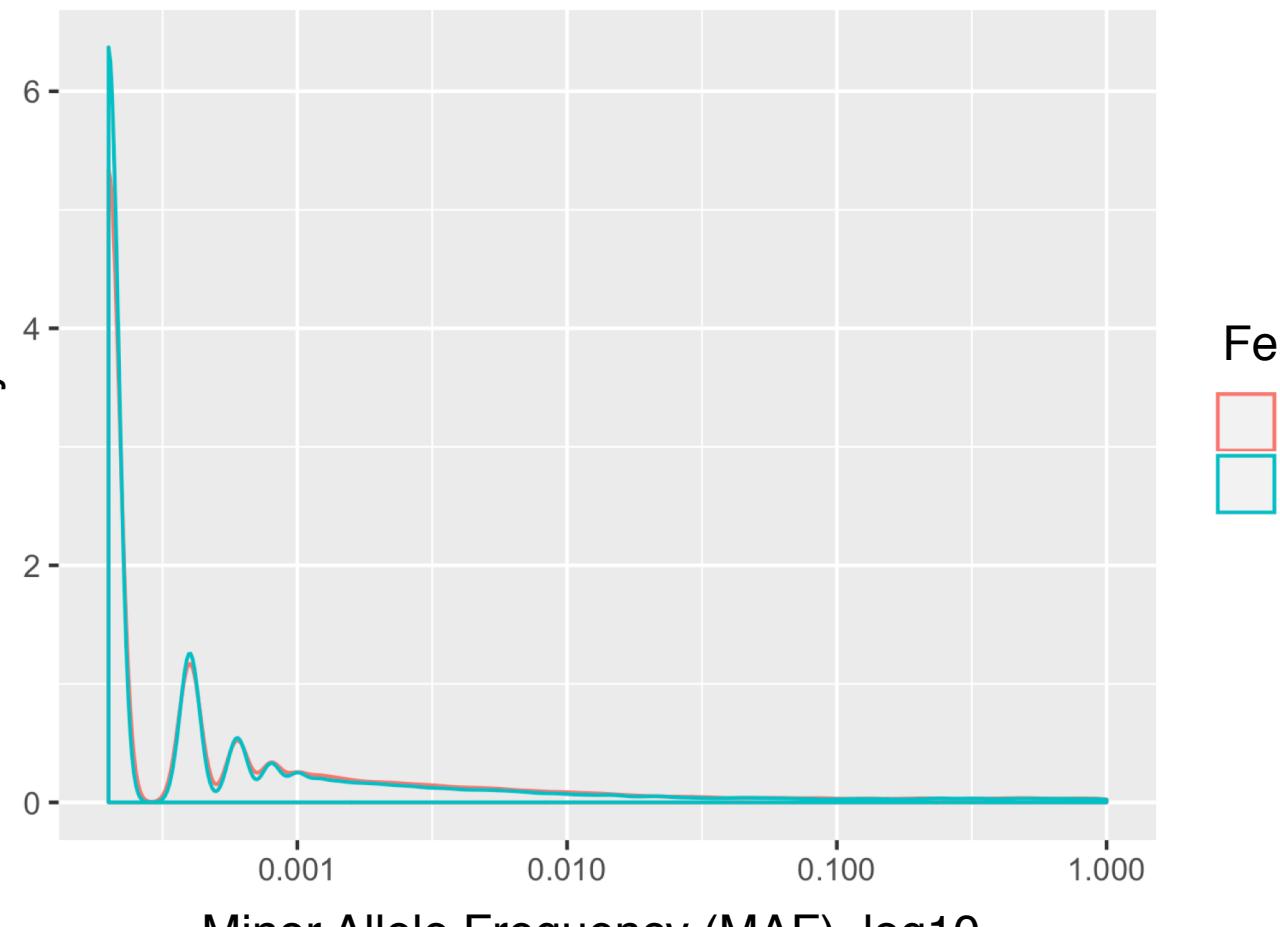
McGarvey et al., found multiple protein domain categories where there was an enrichment for pathogenic ClinVar SNPs.



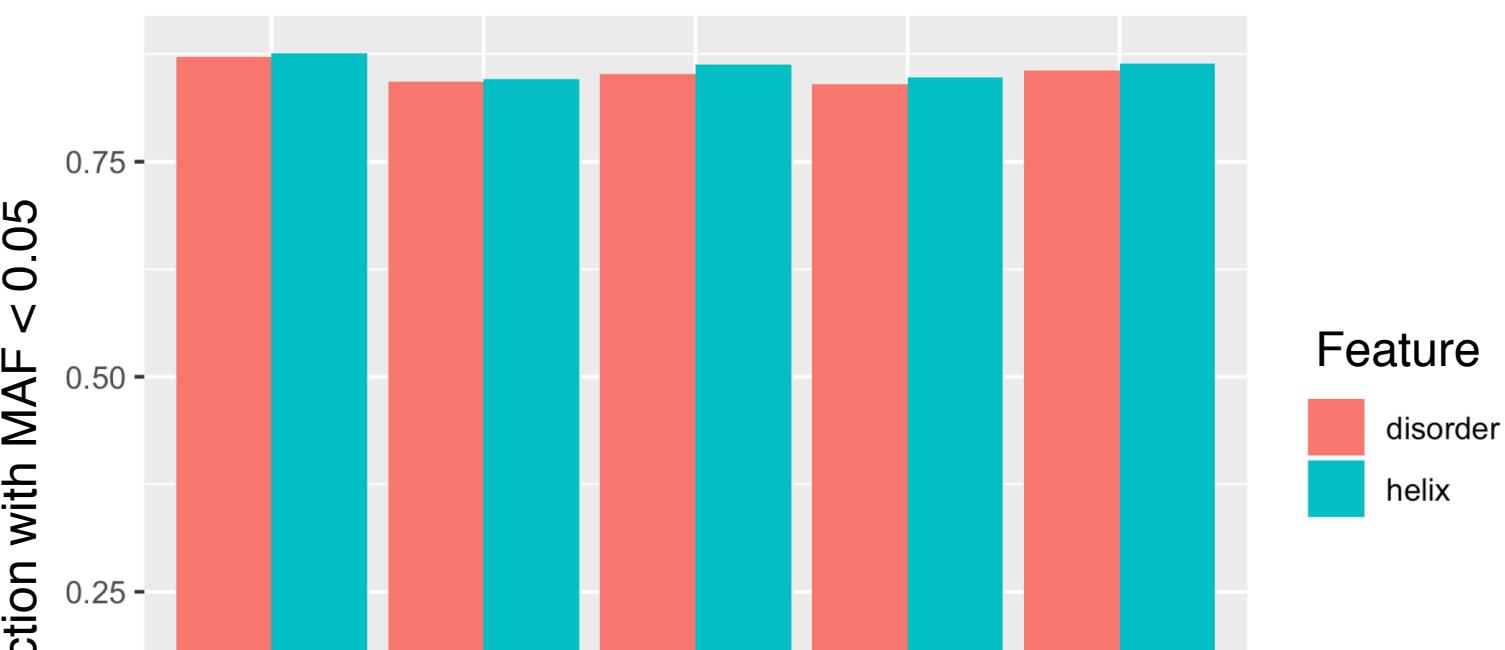
Highlights the utility of investigating within the context of protein features.

## Utility of characterizing genetic variation within protein features

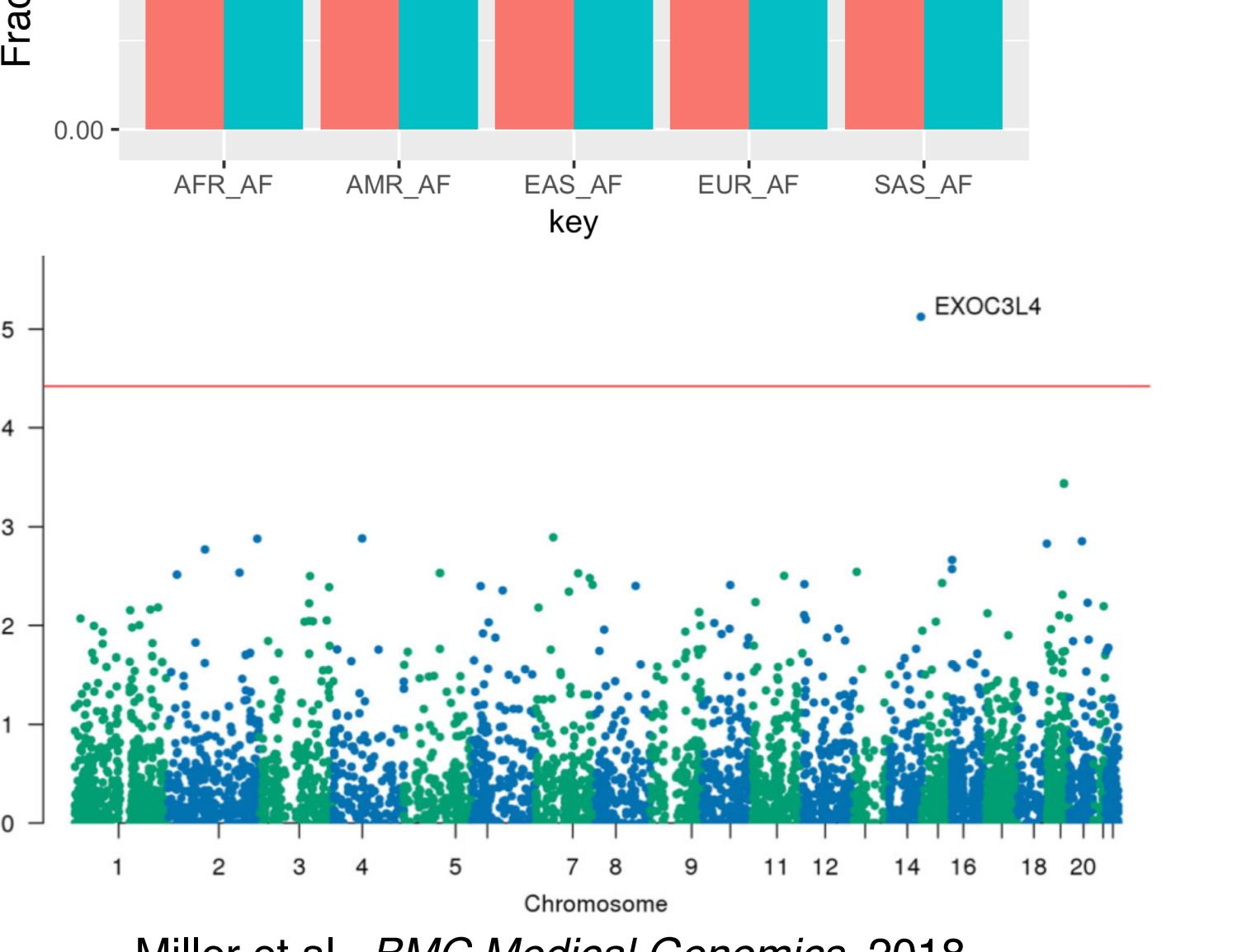
Overall frequency of variants from 1000 genomes project in protein features.



Percent of low variants across different 1000 genomes ancestries and protein features.



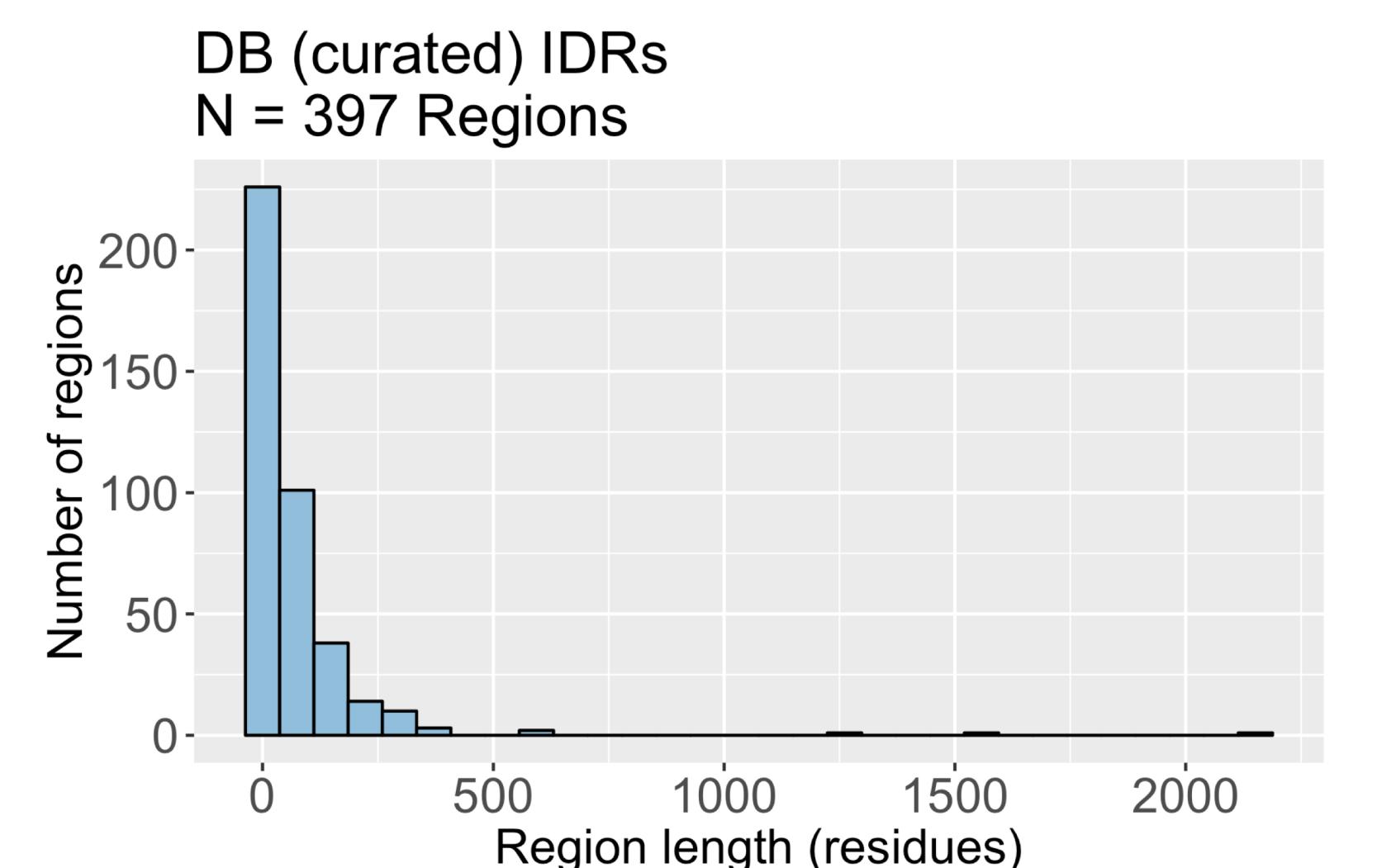
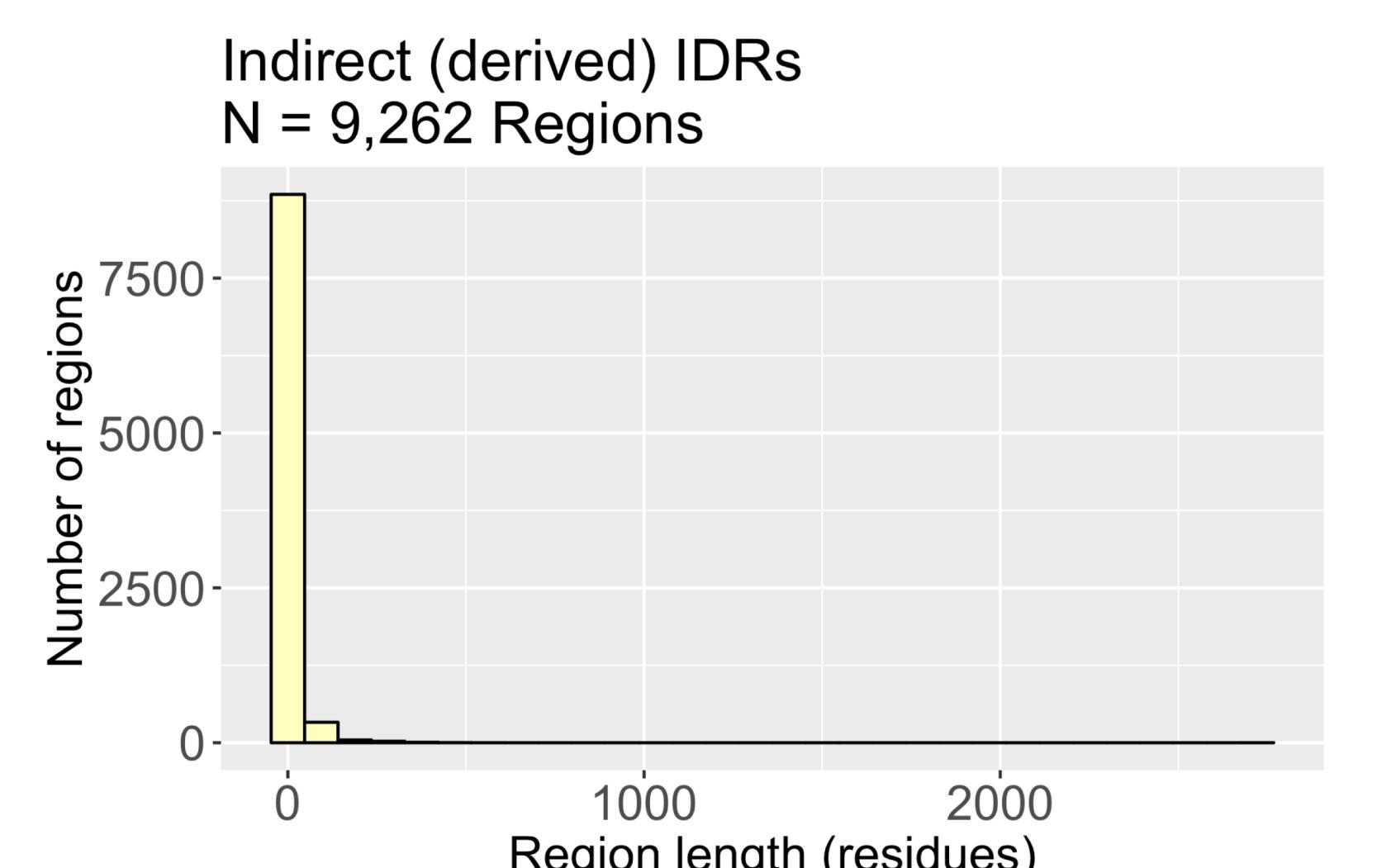
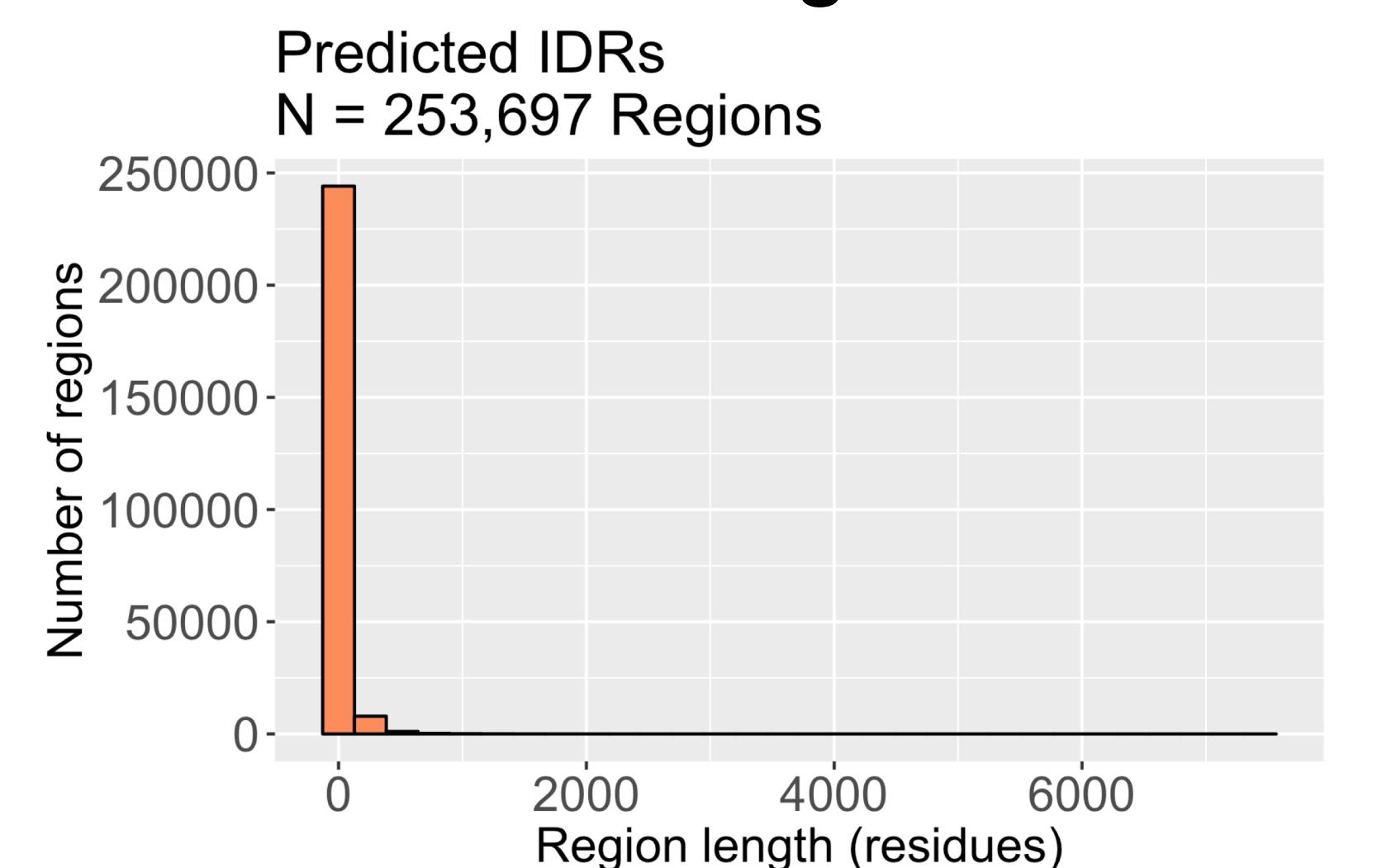
Perform rare variant association tests between AD cases and controls to identify genes associated to AD through SNVs in their IDRs.



Miller et al., BMC Medical Genomics. 2018

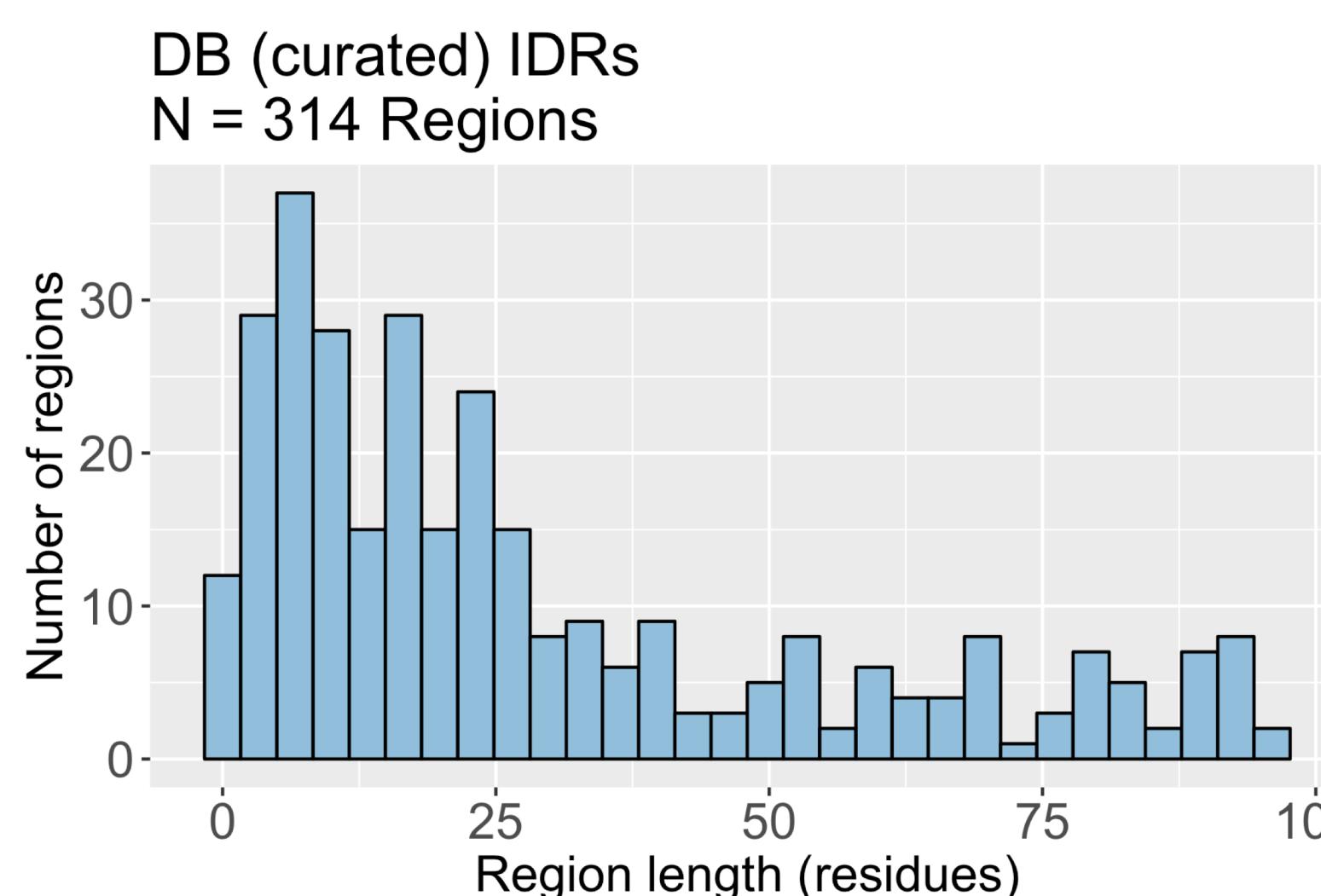
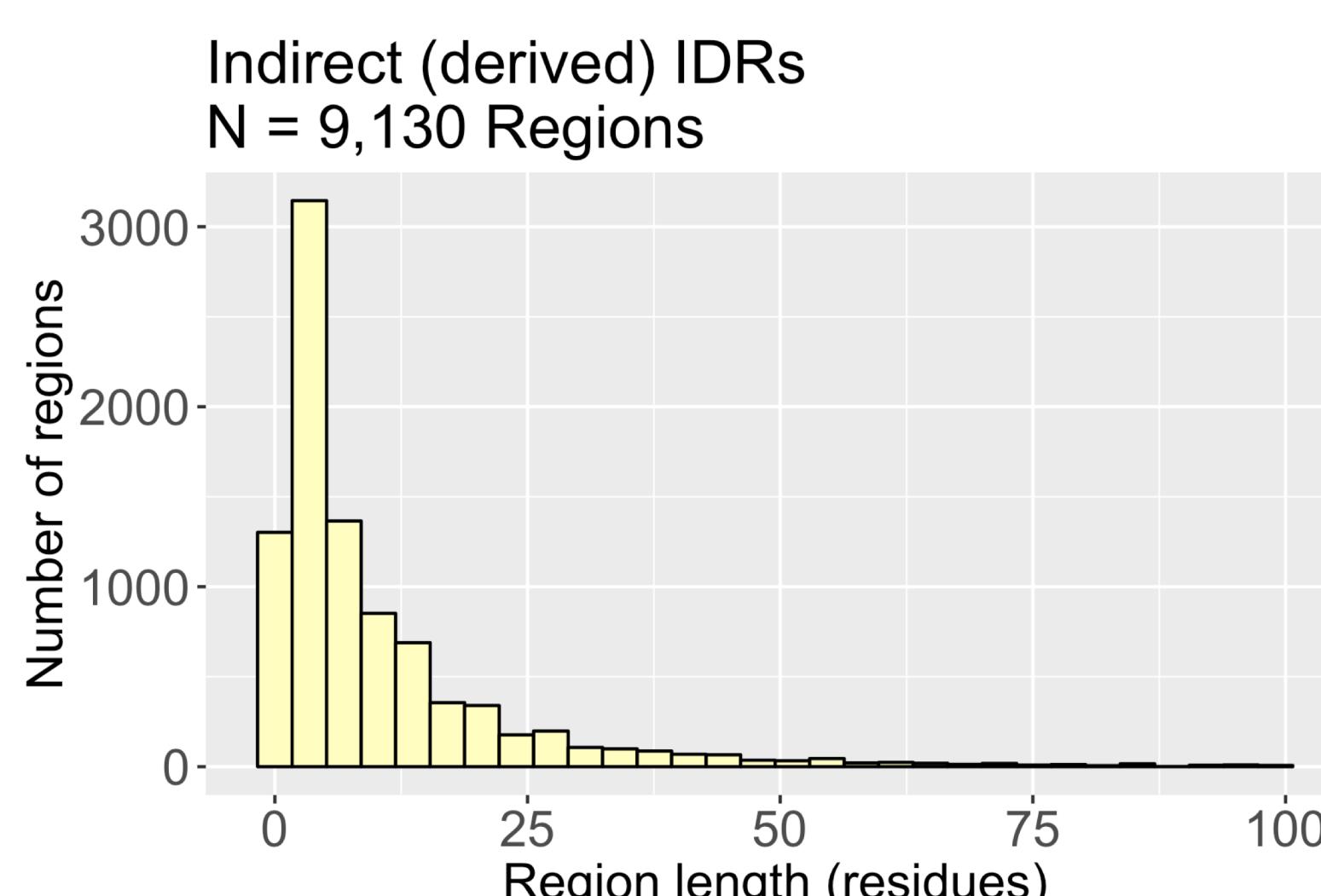
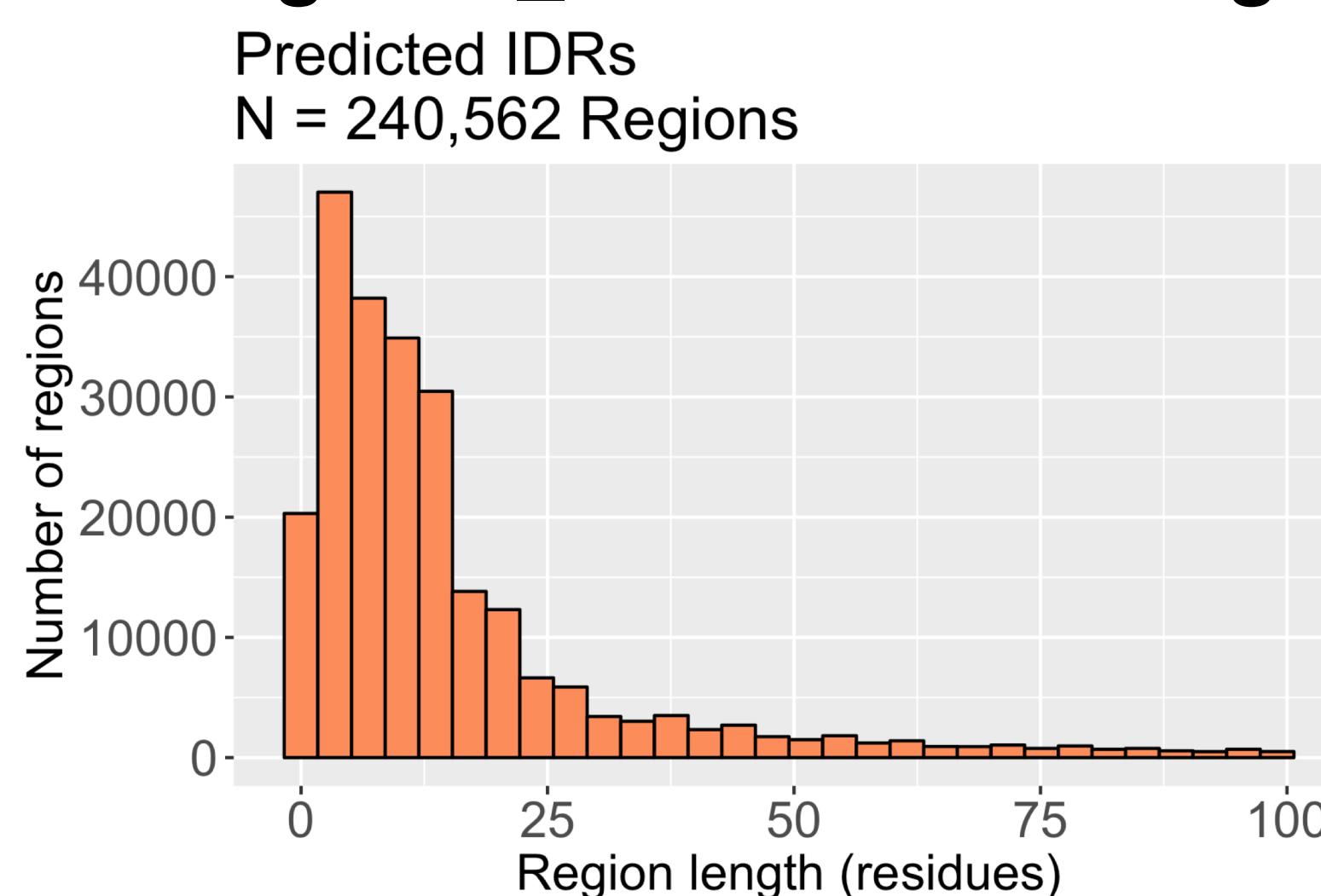
## Characterizing disordered regions using the MobiDB database

### All Regions



Protein IDs were filtered for Swiss-Prot reviewed proteins. All three types of disordered regions have long tails, but are primarily made up of sequences smaller than 100 residues. There are many more predicted regions than derived or curated regions.

### Regions ≤ 100 residues long



## Future Questions

- How does genetic variation differ in known functional regions of IDRs?
- Does the co-occurrence of certain domains impact MAF in IDRs?
- Does variation differ between those with/without AD/MCI?
- How does genetic variation in IDRs differ in other diseases?

## Acknowledgments and links

Thank you to the Ritchie Lab for feedback. Funding source: NHGRI T32HG009495-01.



Poster/GitHub



Twitter @JEMgenes



Lab Website

