Understanding Human Disease through Millions of Tandem Repeats across Millions of Genomes

Wei Li, PhD

Grace B. Bell Endowed Chair and Professor of Bioinformatics

Division of Computational Biomedicine

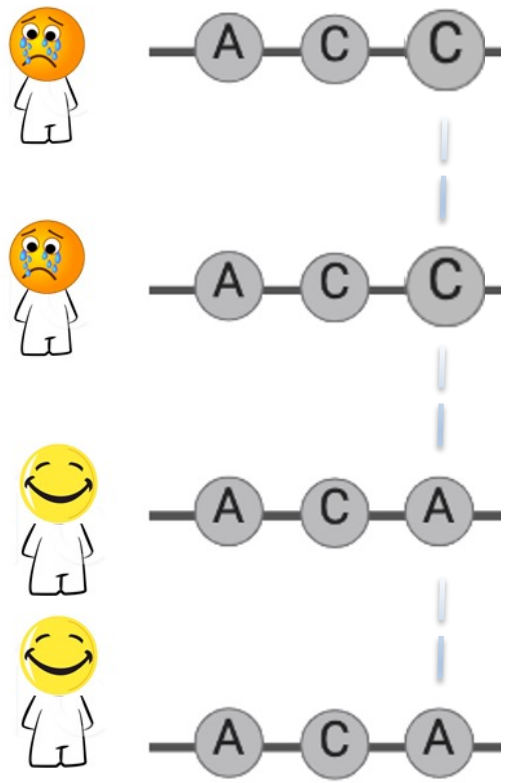
Department of Biological Chemistry

School of Medicine, University of California, Irvine

https://sites.uci.edu/weililab

COI Disclosures: ChosenMed and Helio Genomics

Germline SNPs and Somatic Mutations Associated with Diseases



Challenges in SNP Association Studies

- **Missing Diagnosis**
  - **>60%** of rare genetic diseases remain undiagnosed.
- **Missing Heritability**
  - in common genetics disease like prostate cancer, breast cancer, and diabetes
  - **>50%** patients not explained by GWAS SNPs

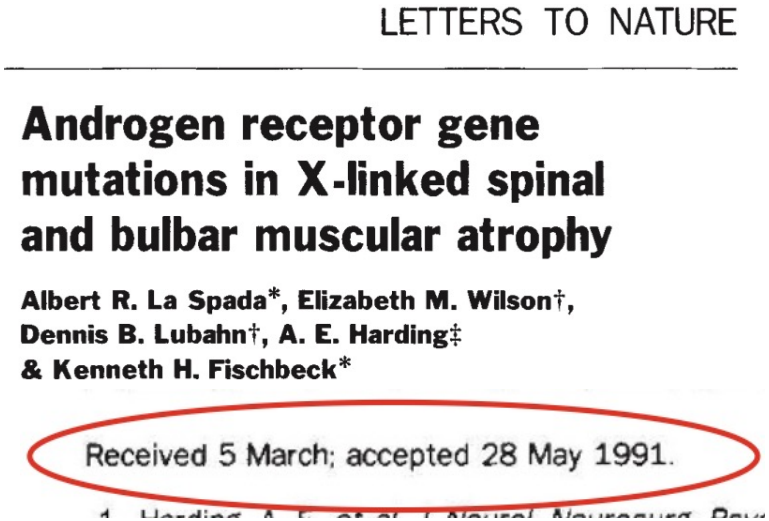
Missing diagnosis and heritability: a major barrier to precision medicine

Hypothesis: Tandem Repeats (TRs) are key genetic factors contributing to undiagnosed diseases and missing heritability

Rationale: beyond the ~60 well-known TRs linked to diseases like Huntington's Disease, ALS, and SBMA, millions more exist in the genome and have been largely **excluded** from clinical testing and GWAS.

Leslie Thompson

Albert La Spada



LETTERS TO NATURE

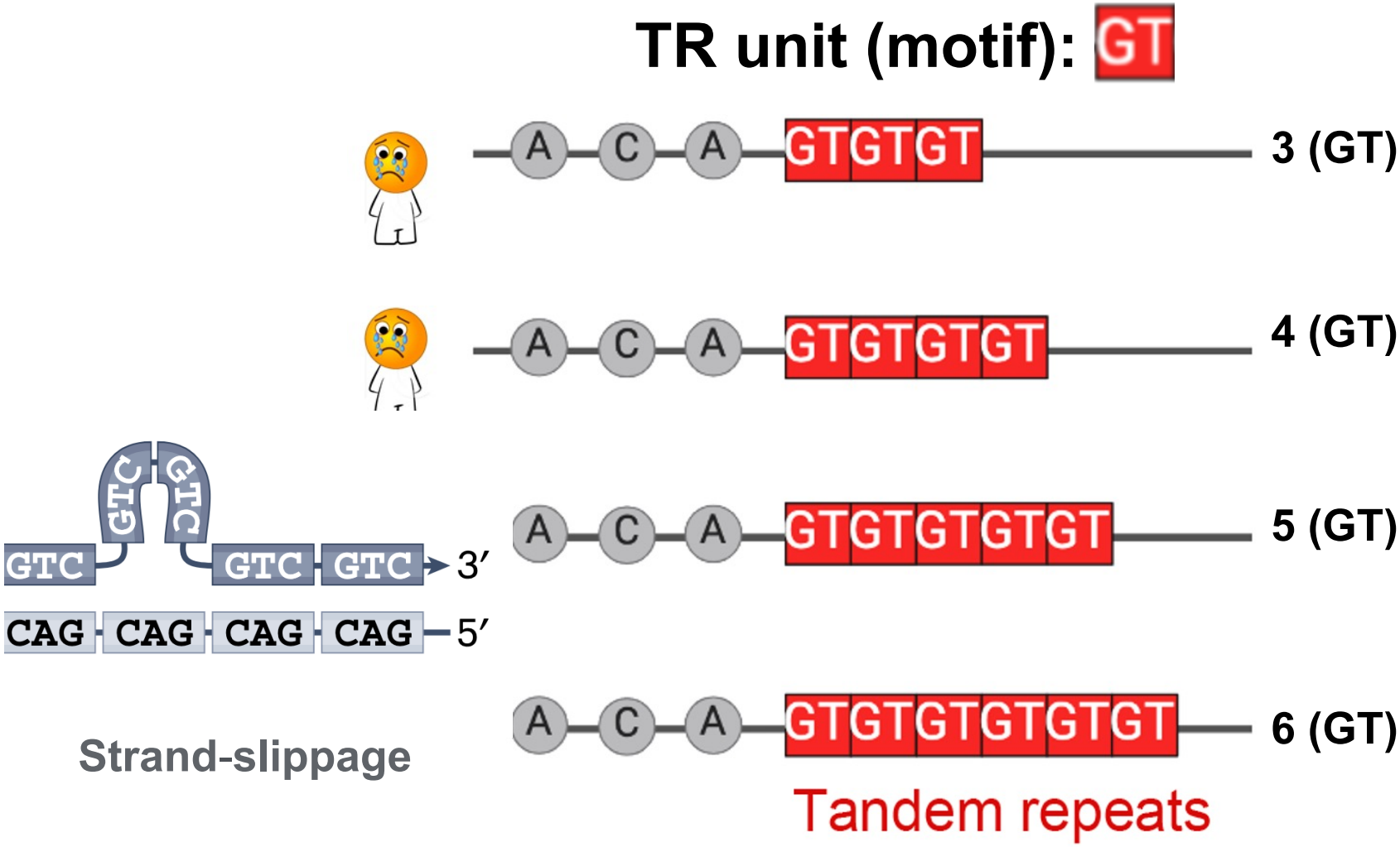
Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy

Albert R. La Spada*, Elizabeth M. Wilson†, Dennis B. Lubahn†, A. E. Harding‡ & Kenneth H. Fischbeck*

Received 5 March; accepted 28 May 1991.

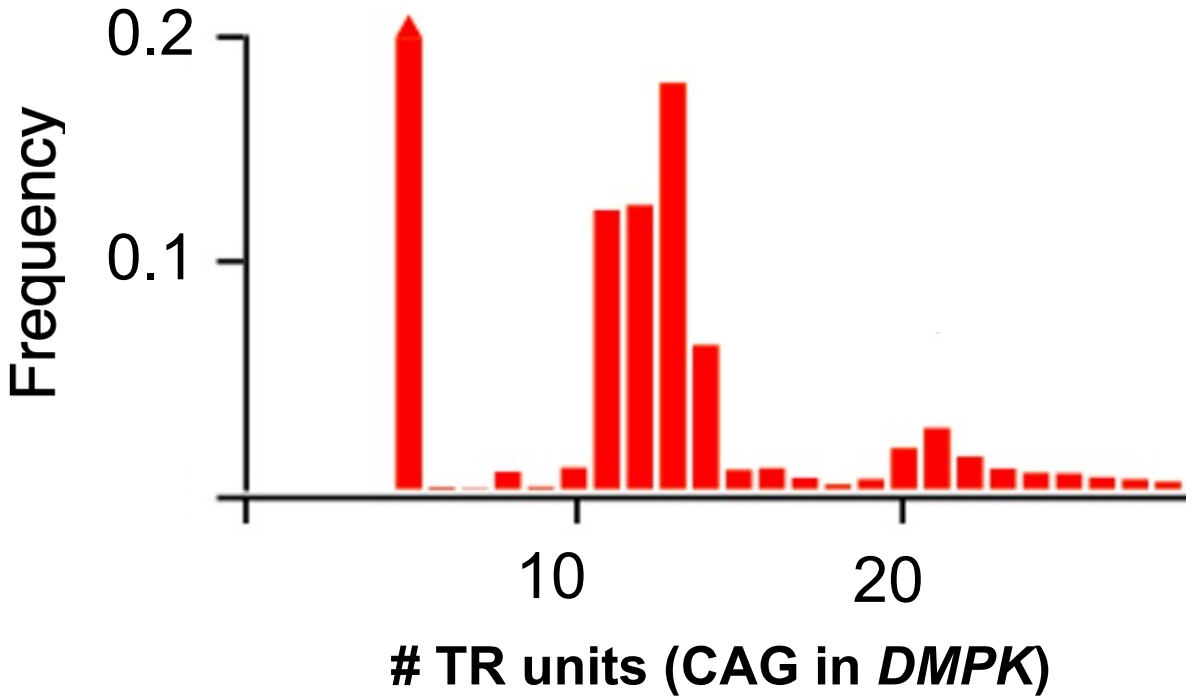
discovered the **FIRST** repeat expansion disorder in summer 1990

<u>Hypothesis</u>: Tandem Repeats (TRs), excluded from clinical testing and GWAS, are crucial genetic factors contributing to missing diagnosis & heritability



Tandem Repeat (TR) Variations

- TRs comprise ~8% of the genome across ~10 millions loci
- TR expansions or contractions generates 2-1,000 of alleles
- TR variation rate is ~1,000 times higher than SNPs
  - 1e-5 vs 1e-8 per locus per generation
- TRs are hard to impute with SNPs (A/T) and require whole genome sequencing

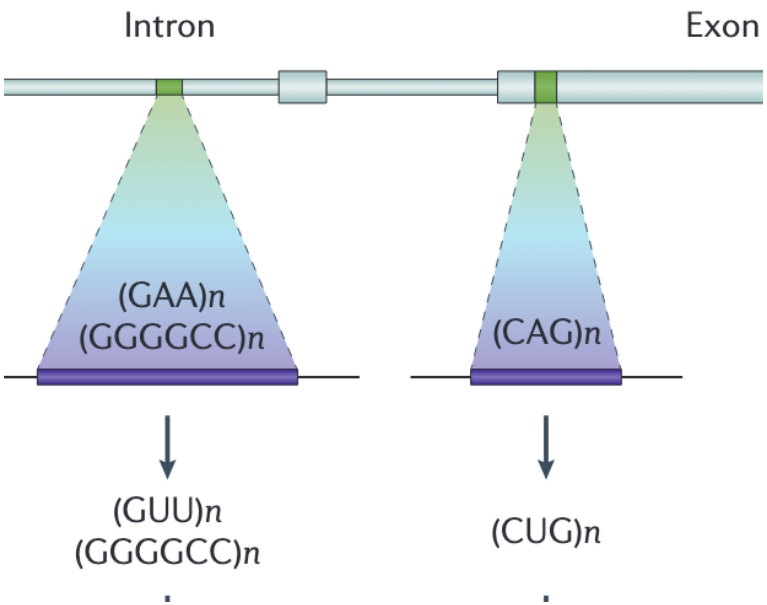


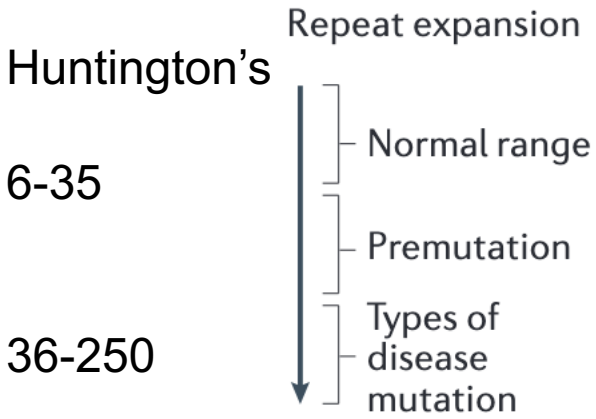
Knowledge Gaps: Tandem Repeats have been Implicated in only ~60/6,000 (1%) of Rare Genetic Diseases

primarily through family-based linkage analyses

Intronic TRDs
- FRDA
- C9ORF72 TRDs (includes subset of ALS and FTD)

Polyglutamine TRDs
- HD
- SCA1, SCA2, SCA3, SCA6, SCA7 and SCA17
- SBMA (Kennedy disease)
- DRPLA



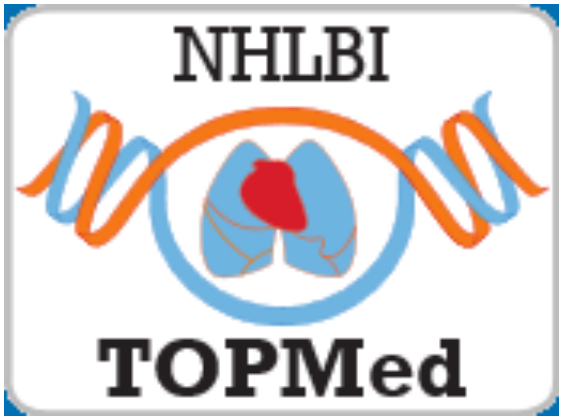


Hannan AJ, 2018

Whole Genome Sequencing (WGS) of Millions of Human

UK

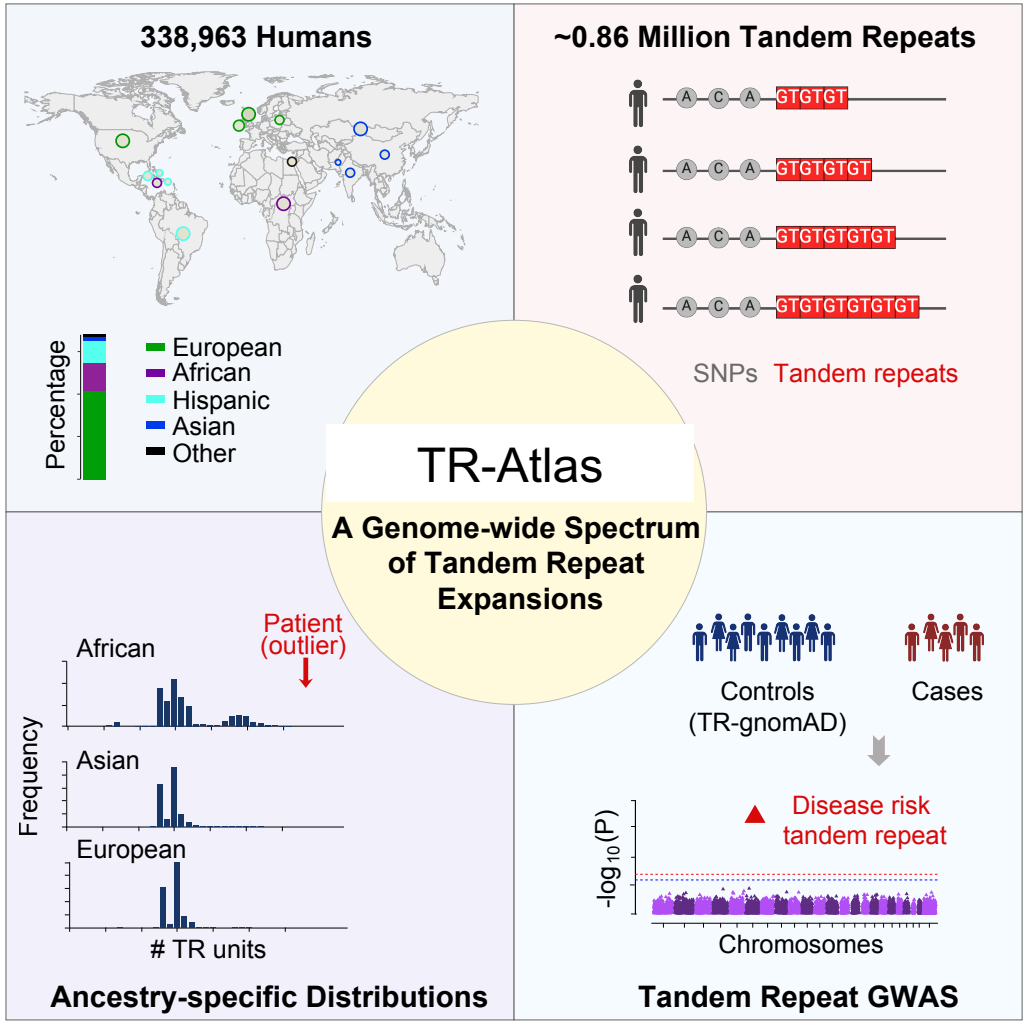
NIH



- Google/Amazon Cloud computing

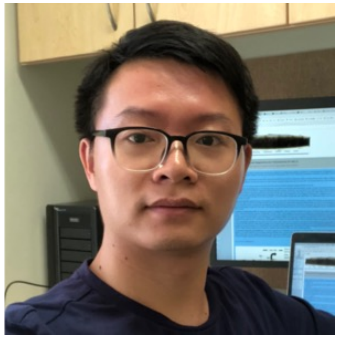
TR-Atlas: ~1 Million TR loci across 340K individuals -- built for $400K (Not $12M)

-- 100 times more than WebSTR



https://wlcb.oit.uci.edu/TRatlas/

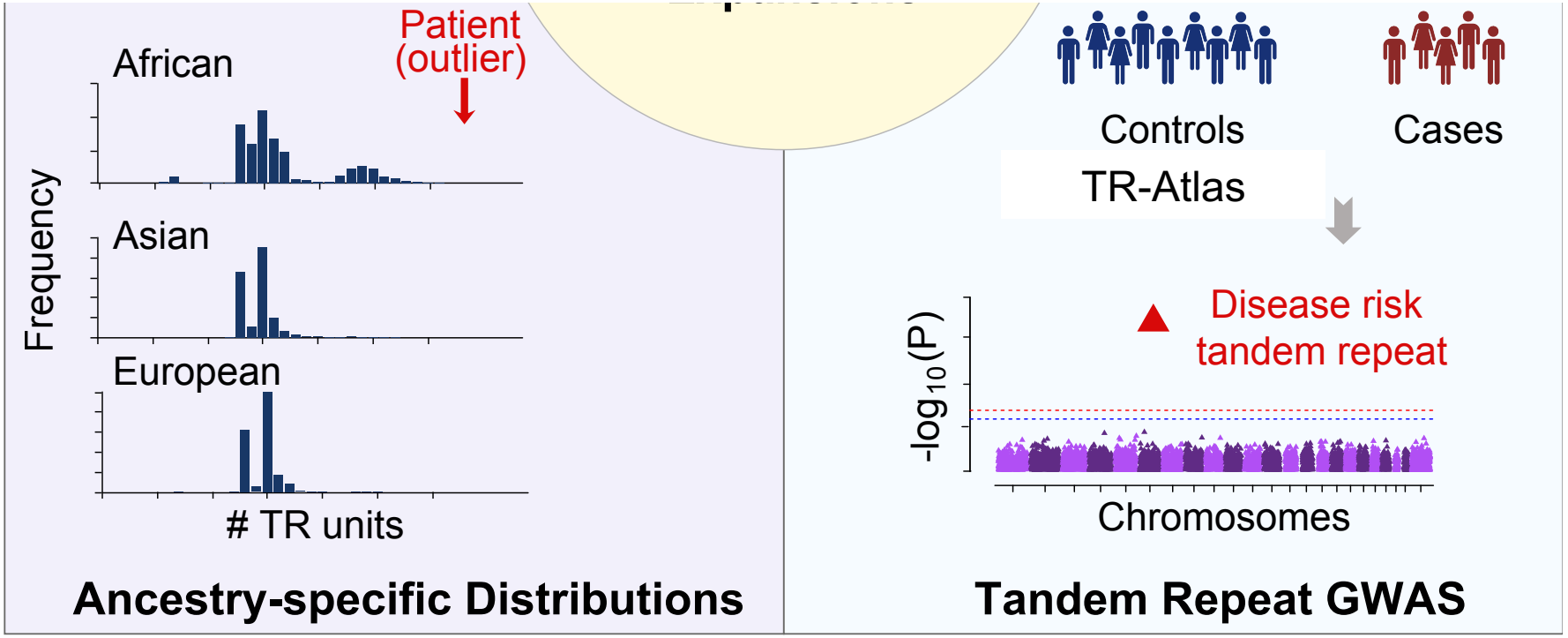
Short read WGS with TR genotypers: ExpansionHunter, GangSTR



Ya Allen Cui

Cui *et al.*, ***Cell*** *2024*

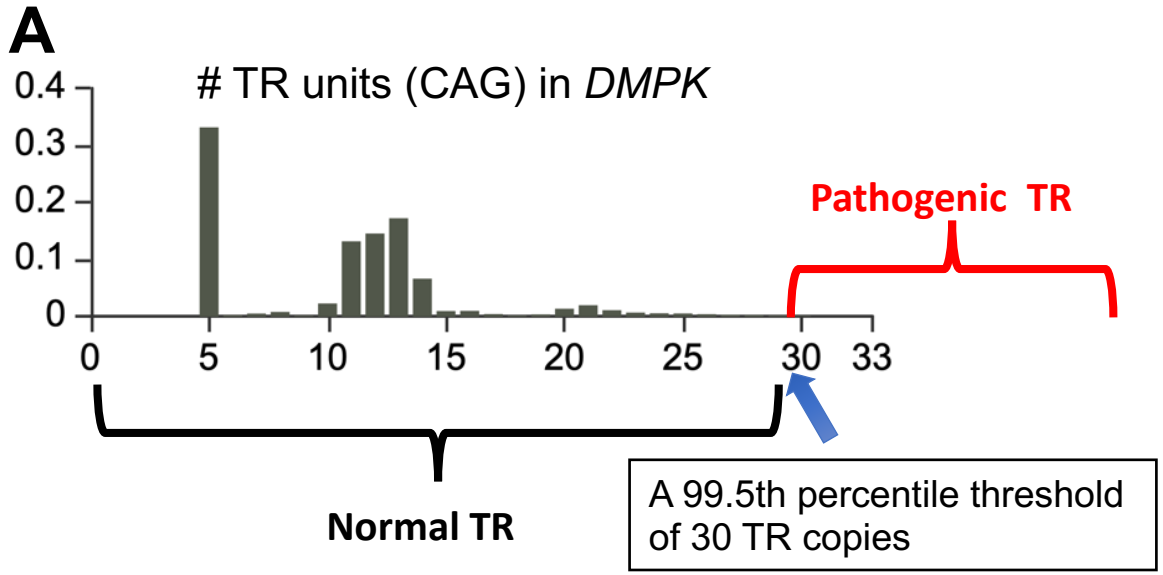
TR-Atlas



Cui et al., ***Cell*** *2024*

Threshold-Based Evaluation of TR Pathogenicity

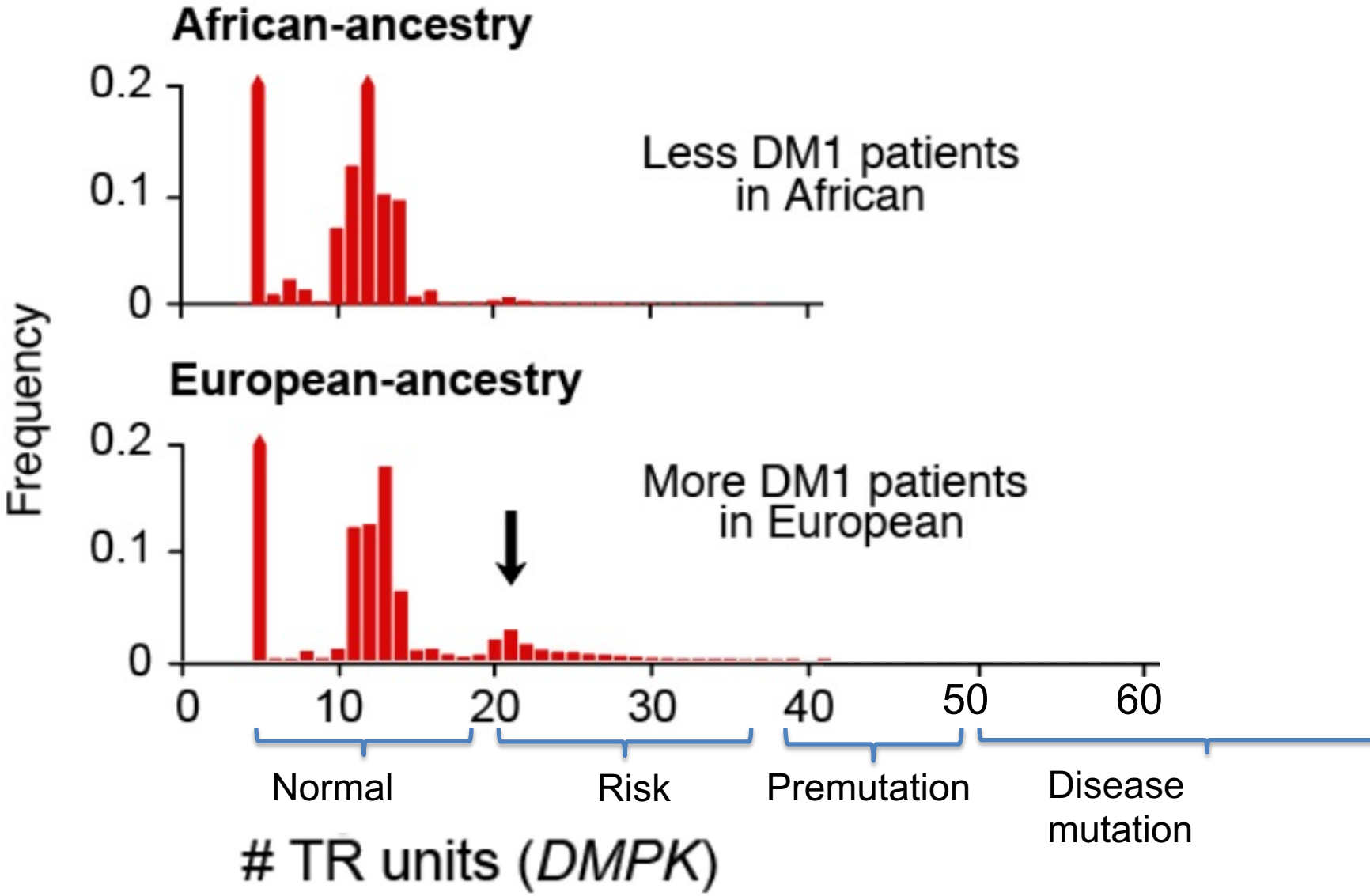
does not require precise allele length measurements



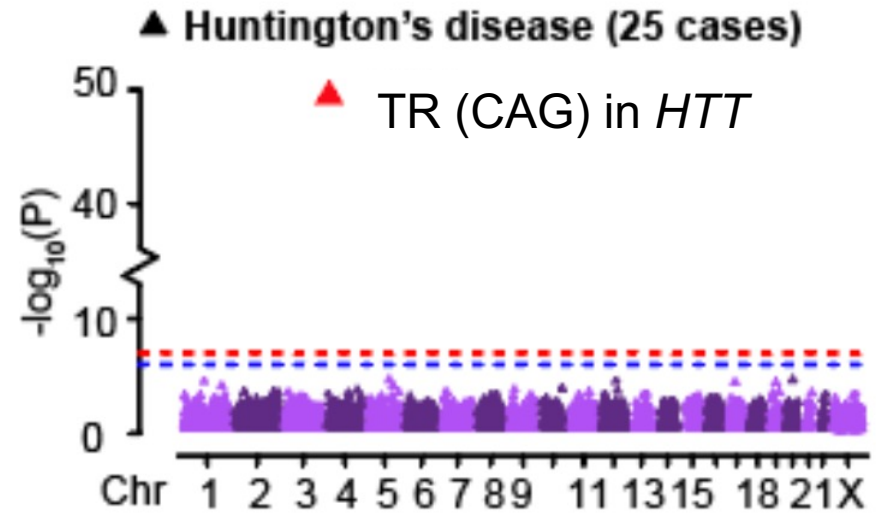
- short-read WGS with ExpansionHunter can successfully separate normal and risk TR groups **with 97.3% sensitivity and 99.6% specificity** in 404 patients across 13 disease-associated TR loci (including 4 loci with large repeat expansions) when compared to the gold-standard diagnostic method, PCR testing

THE LANCET Neurology 2022

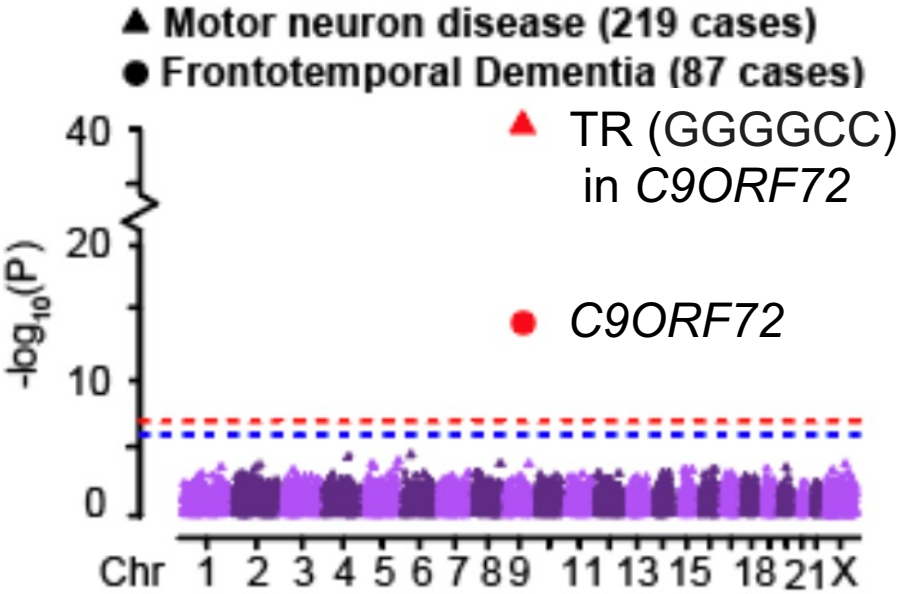
Tandem Repeat Expansions Explain Health Disparities



Tandem Repeat -based genome-wide association studies (TR GWAS)



vs. 18,000 healthy controls in TR-Atlas with matched ancestry



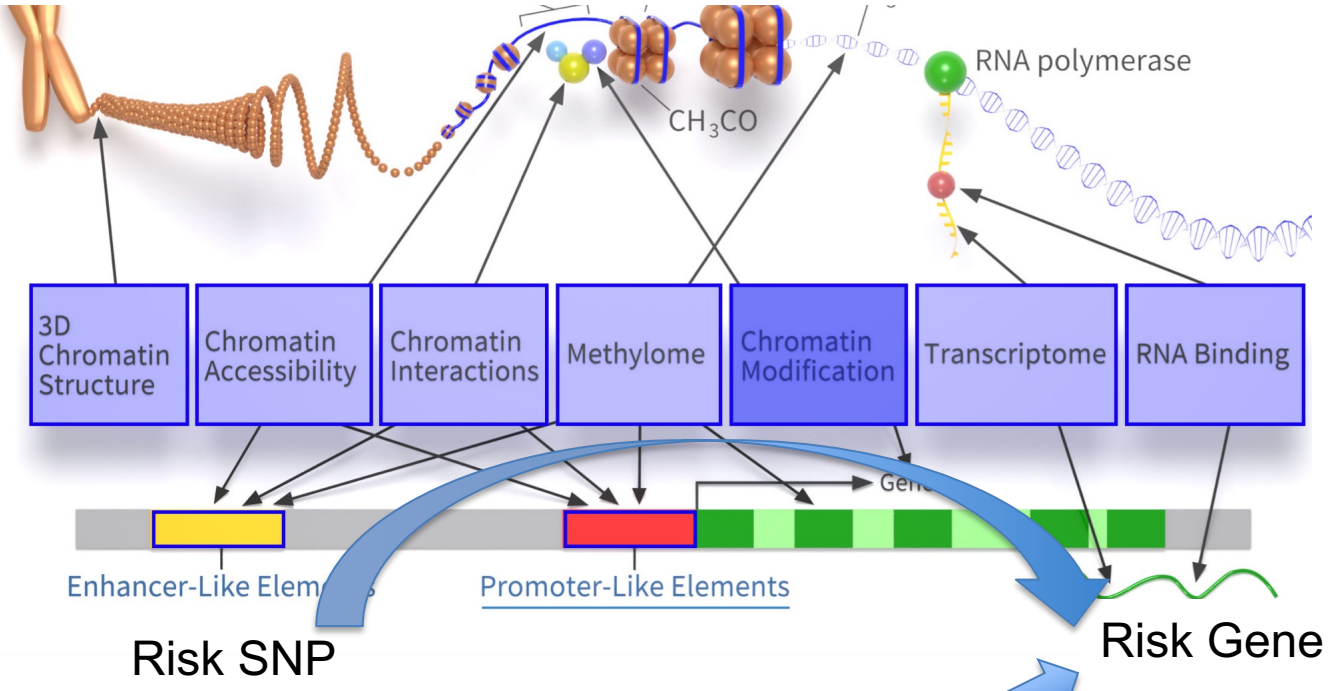
vs. 18,000 healthy controls in TR-Atlas with matched ancestry

Challenges in SNP Association Studies

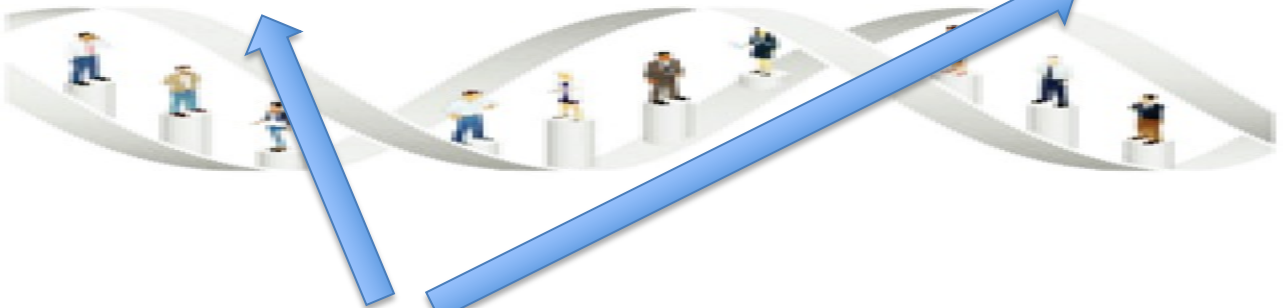
- **Missing Diagnosis**
  - ✓ **>60%** of rare genetic diseases remain undiagnosed.
- **Missing Heritability**
  - ✓ Genetics explain >**60%** of common diseases (e.g., prostate/breast cancer, diabetes).
  - ✓ **>50%** patients not explained by GWAS SNPs
- **Missing Mechanisms for Non-coding Variants (>95% total)**

eQTLs: SNPs associated with gene regulation

Molecular Phenotype



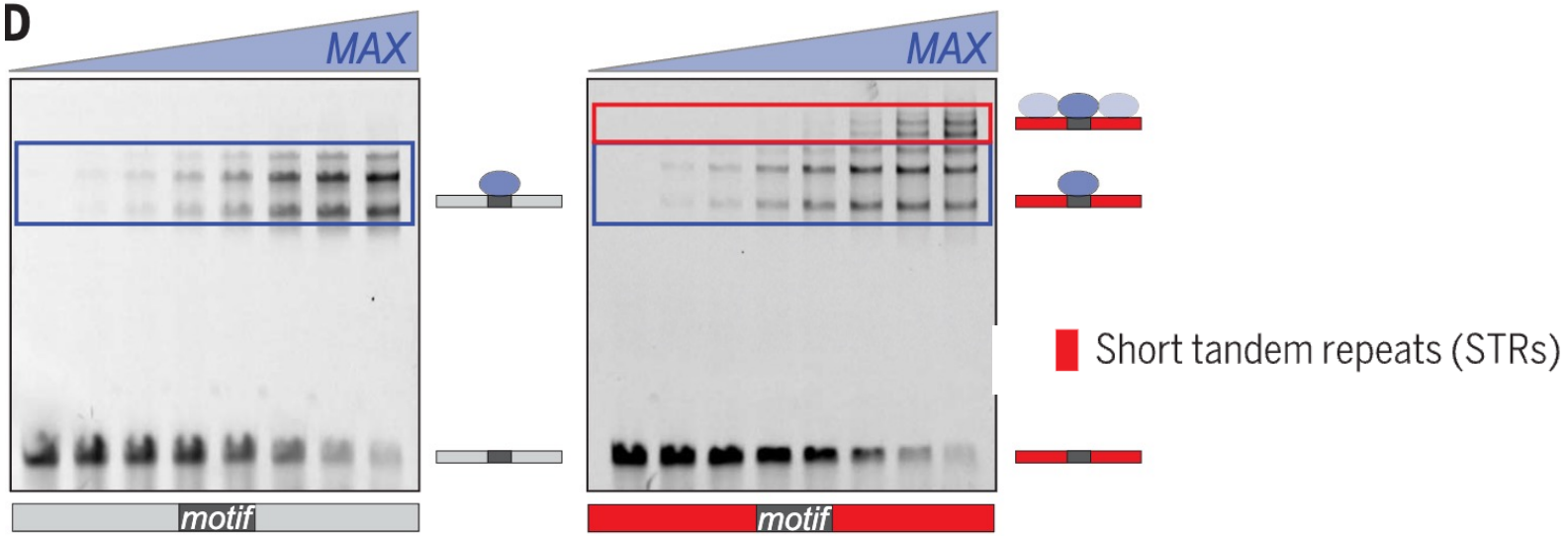
Genotype



Phenotype

Short tandem repeats bind transcription factors to tune eukaryotic gene expression

Horton et al., ***Science*** 2023



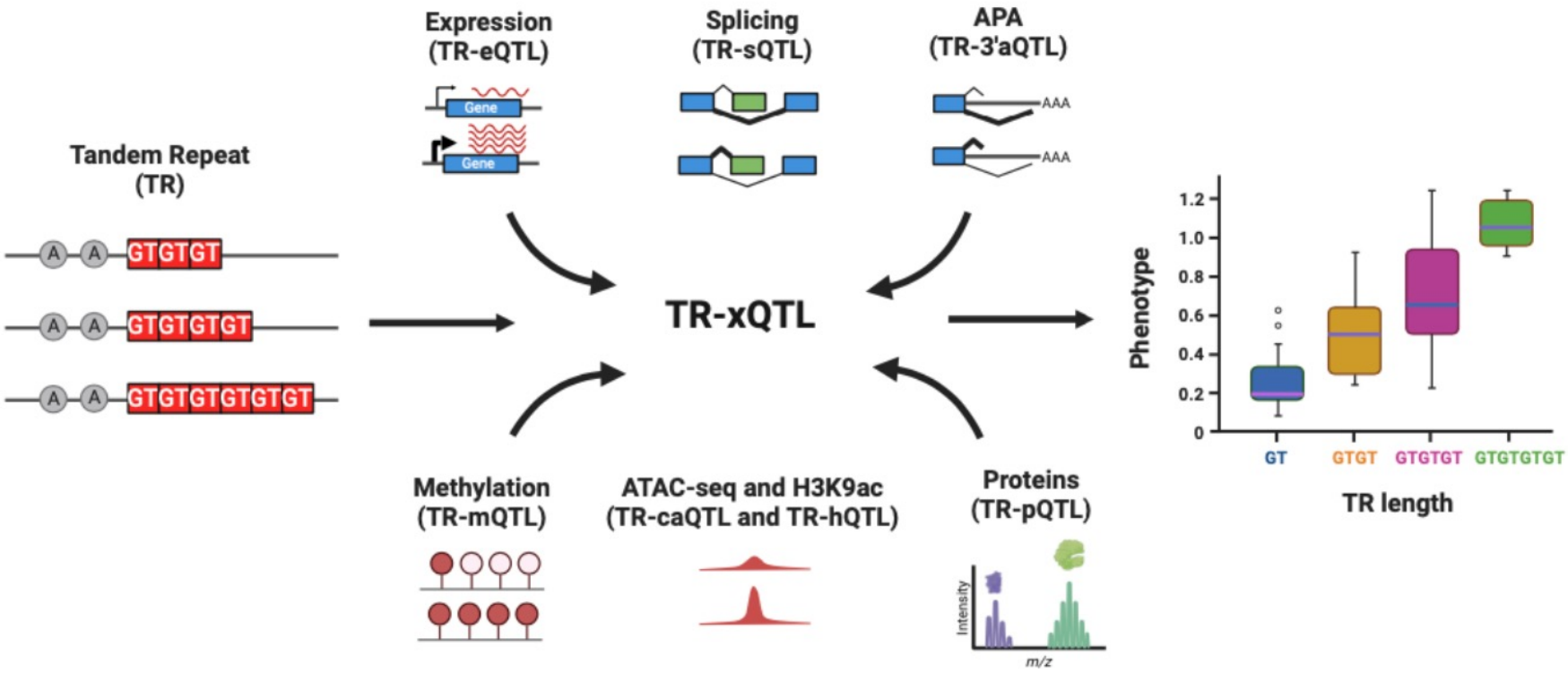
Analysis | Published: 01 November 2019

The impact of short tandem repeat variation on gene expression

Stephanie Feupe Fotsing, Jonathan Margoliash, Catherine Wang, Shubham Saini, Richard Yanicky, Sharona Shleizer-Burko, Alon Goren & Melissa Gymrek

Non-coding TR variations affect gene regulation, thereby mediating disease risk

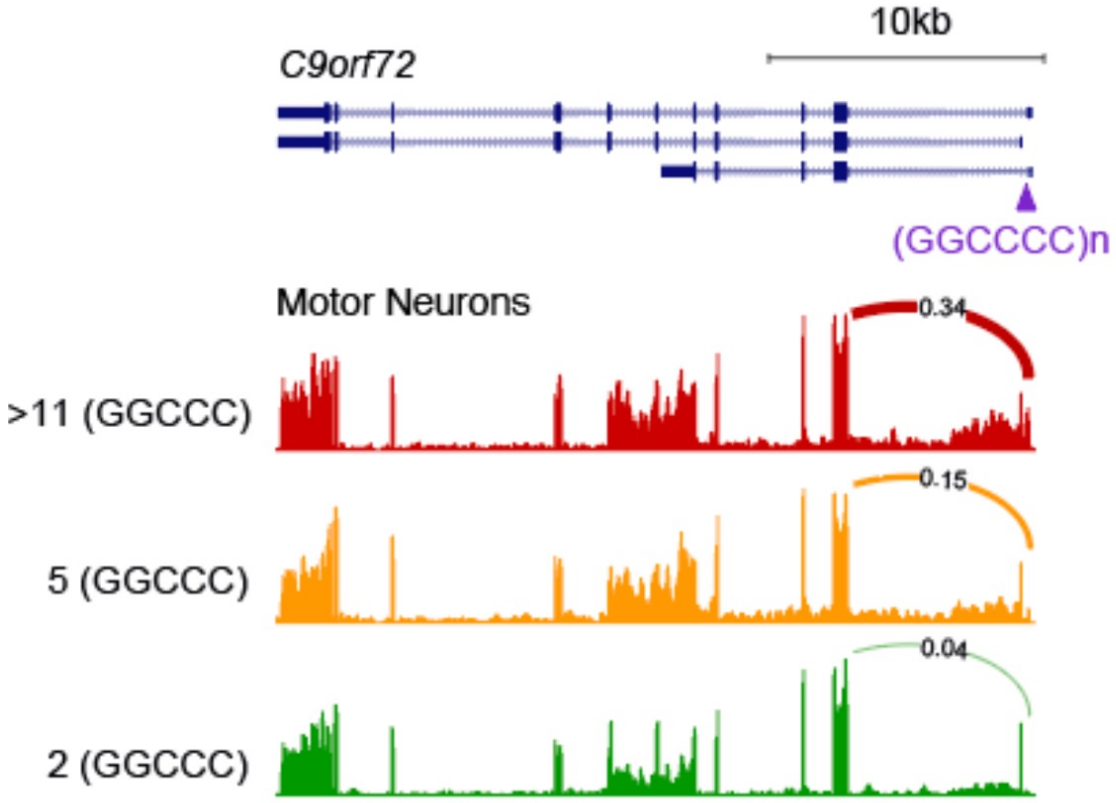
- TR quantitative trait loci (TR-xQTL) across ~4,000 multi-omics samples derived from ~1,600 unique donors
- ~2.2M TR-xQTLs linking ~139K unique TRs to gene regulation

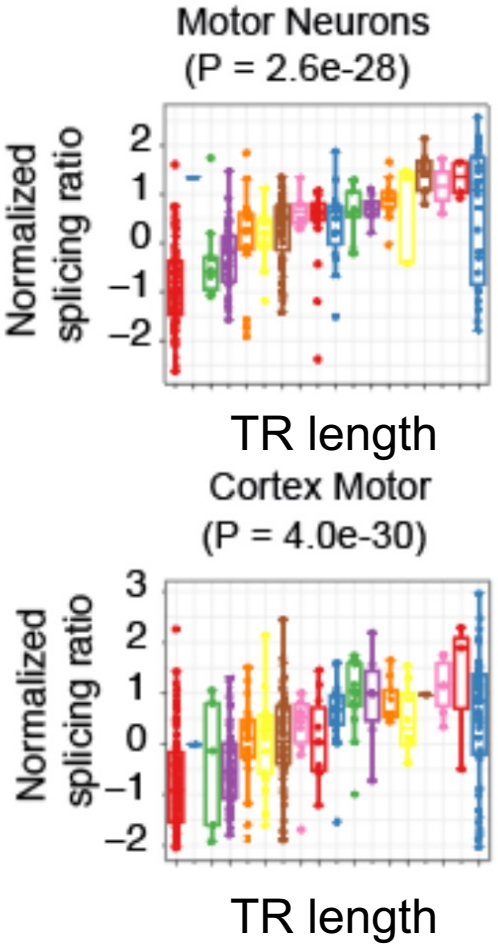


https://wlcb.oit.uci.edu/TRxQTL/

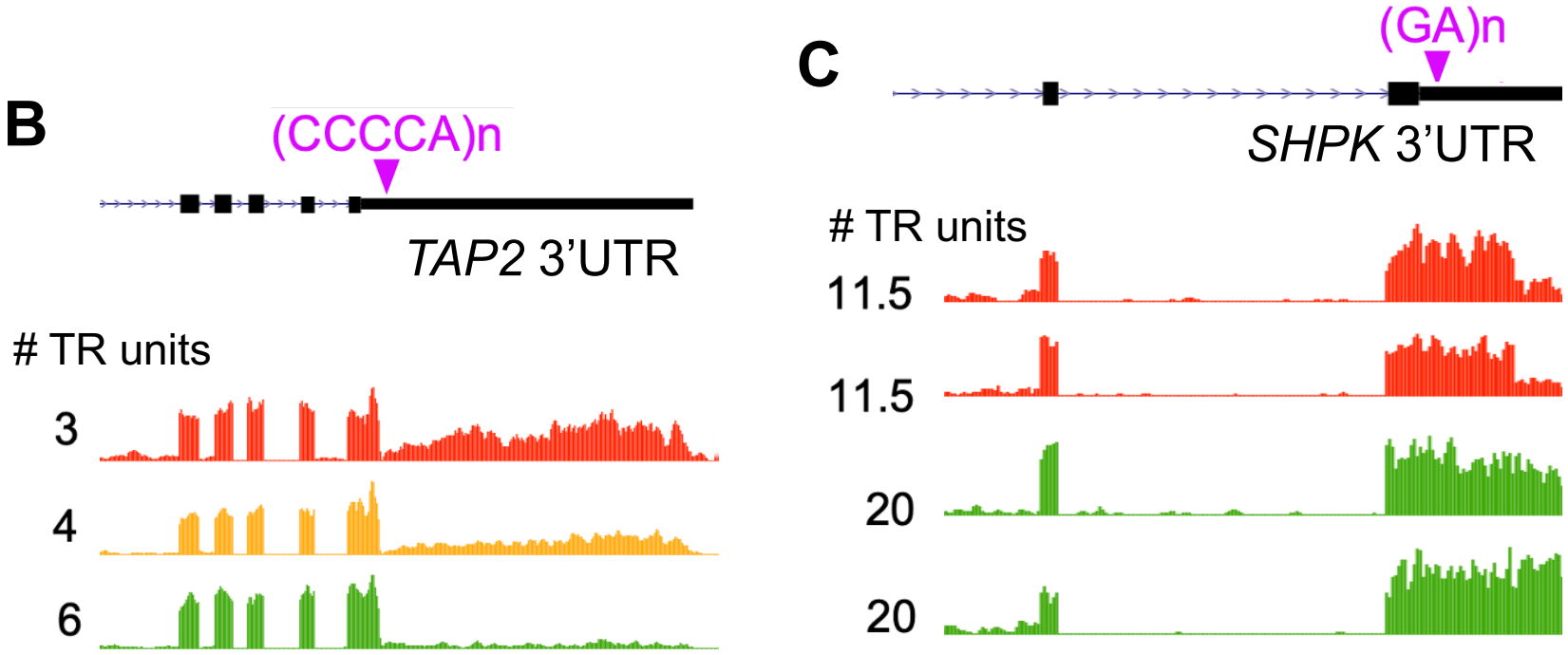
Cui et al., ***Nature Genetics,*** 2025

TR-sQTL of *C9orf72* in ALS



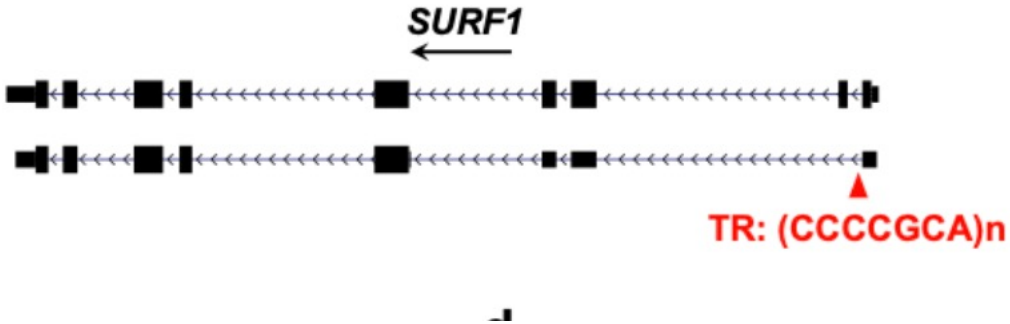


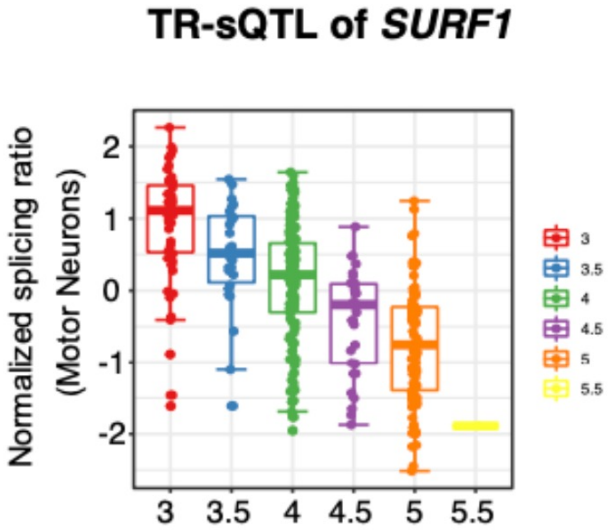
TR 3'UTR APA Quantitative Trait Loci (TR-3'aQTL):

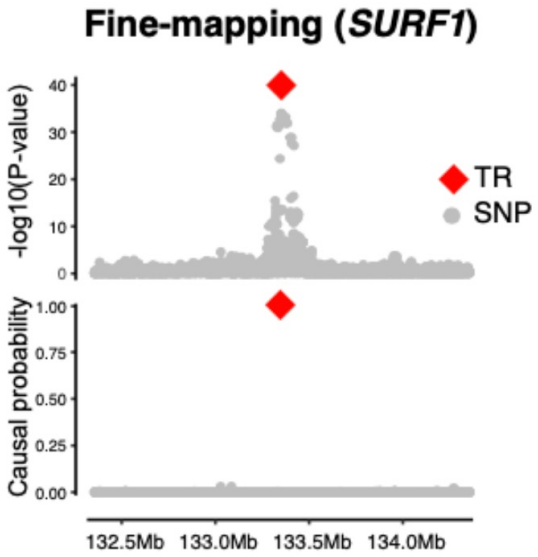


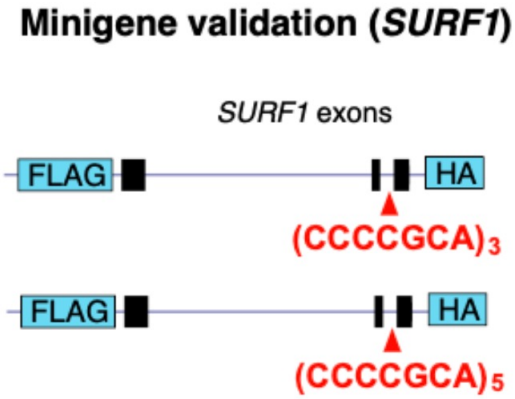
TRs are causal variants for gene regulation

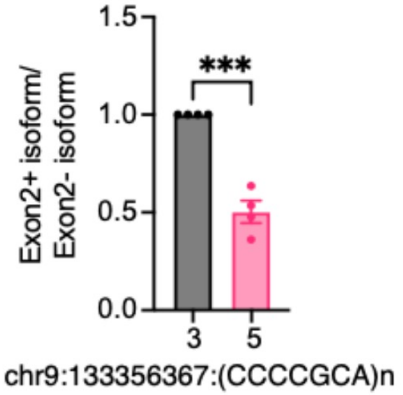
~13% of patients with Leigh syndrome





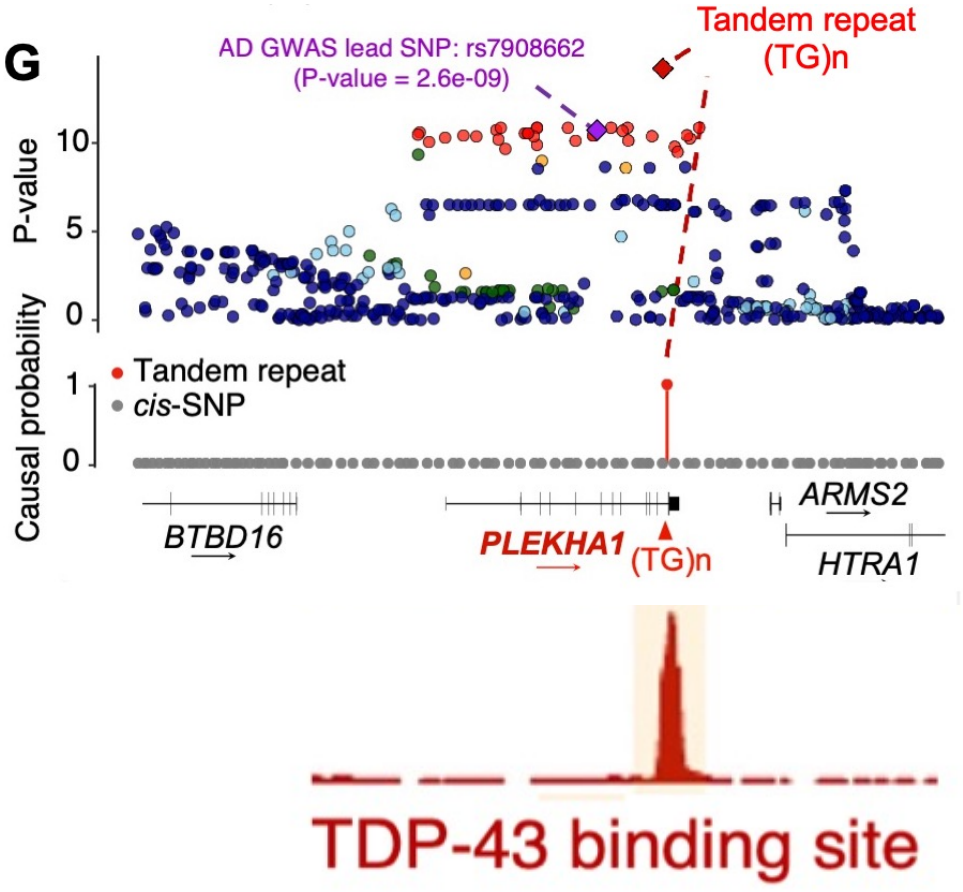


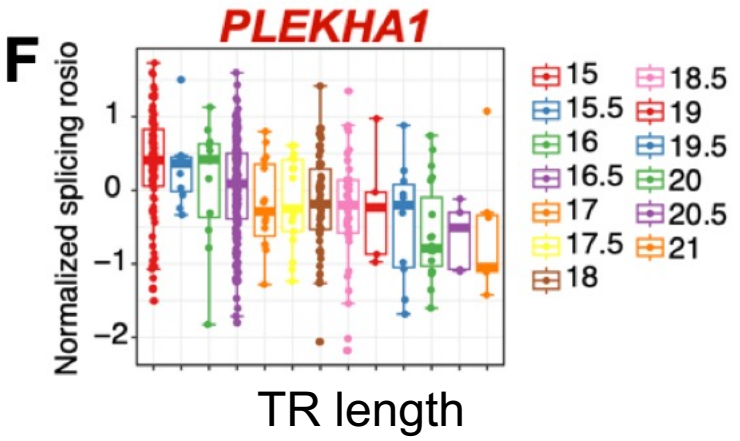


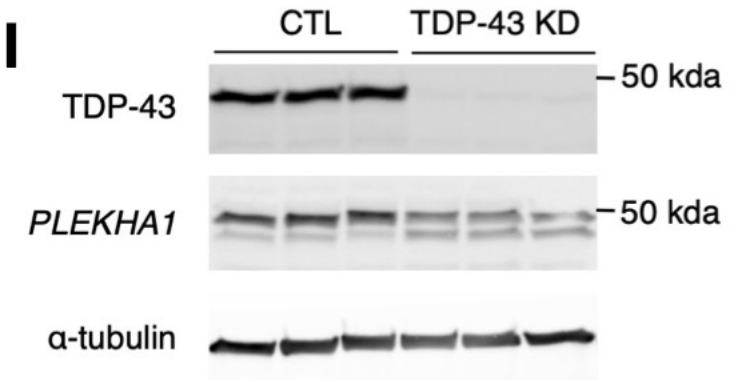


Albert La Spada

A potential risk TR for Alzheimer's disease (AD), primarily by altering TDP-43 binding affinity, leading to alternative splicing of *PLEKHA1*







Shift the focus from SNPs to the broader impact of TR variations

- **TR-Atlas**: ~1 Million TR loci across ~340K Humans (***Cell*** 2024), the world's largest reference map for TR variations
  - TR-GWAS: 2M genomes w/ 8,000 phenotypes: breast cancer, AD, diabetes, ALS
  - TR-outlier: 100K rare disease genomes
- **TR-xQTL**: how TR variants affect gene regulation, thereby mediating disease risk (***Nature Genetics*** 2025)
  - Genetically defined cis-elements for gene regulation beyond traditionally, biochemically defined promoters and enhancers
  - A single gene can be regulated by up to 10 distinct TR loci
- **Call to Collaborate** to unlock the TR's full potential
  - Experimental Biologists, Clinicians, Genomics Labs

Acknowledgements

Wei Li Lab Members

Ya (Allen) Cui
Wenbin Ye
Jason Sheng Li
Chaorong Chen
Xueyi Teng
Zhuoxing Wu
Tanye Wen

Alumni

Liguo Wang (Mayo); Deqiang Sun (Texas A&M);
Kaifu Chen (HMS); Fred Xi (UT Health);
Hyun Park (UPMC); Zheng Xia (Oregon Health);
Jianzhong Su (WMU); Lei Li (SBL); Jiejun Shi (Tongji);
Jianfeng Xu (Helio); Yumei Li (SZU);
Xueqiu Lin (Fredhutch)

Collaborators

Albert La Spada (UCI)
Leslie Thompson (UCI)
Qing Yang (UCI)
Hoda Anto-Culver (UCI)
Eric Vilain (UCI)
Jessica Li (UCLA)
Eric Wagner (Rochester)
Qianben Wang (Duke)

NIH: R01HG007538
R01CA193466
R01CA228140
U54CA217297