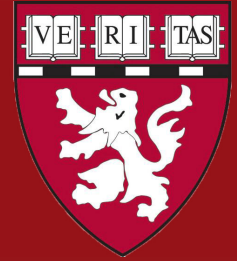




# Quantification of CAG Repeat Instability in Huntington's Disease



**Ricardo Mouro Pinto, PhD**

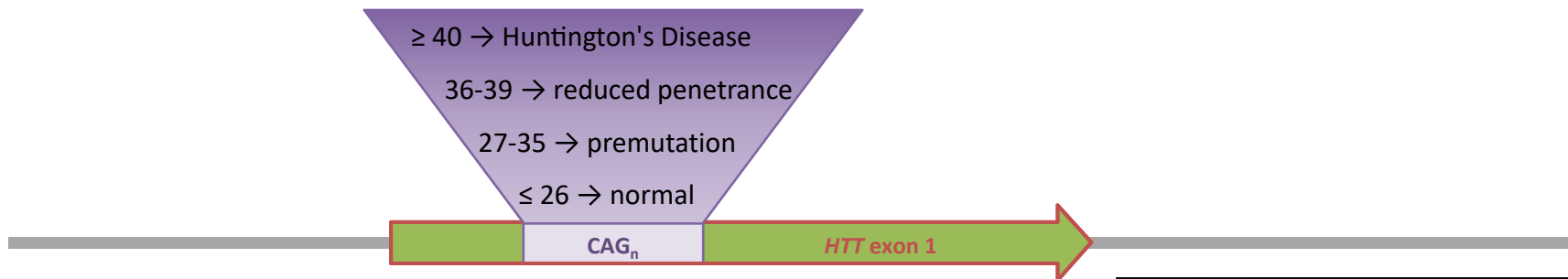
Instructor in Neurology



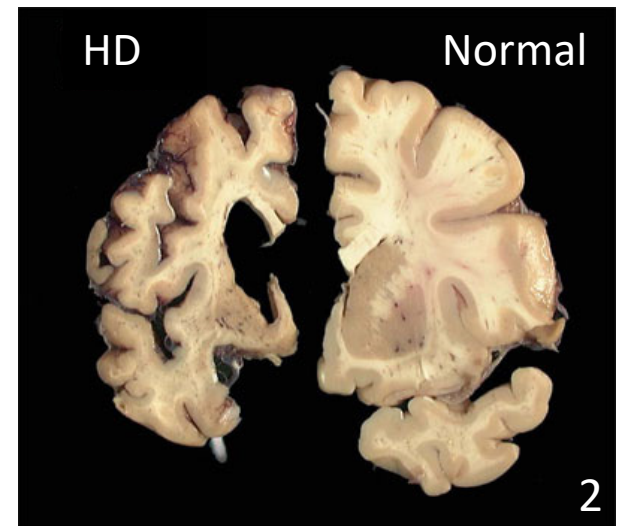
Massachusetts General Hospital  
Harvard Medical School

# Huntington's disease (HD)

- HD is a neurodegenerative disorder caused by the expansion of a CAG repeat tract in exon1 of the *HTT* gene<sup>1</sup>

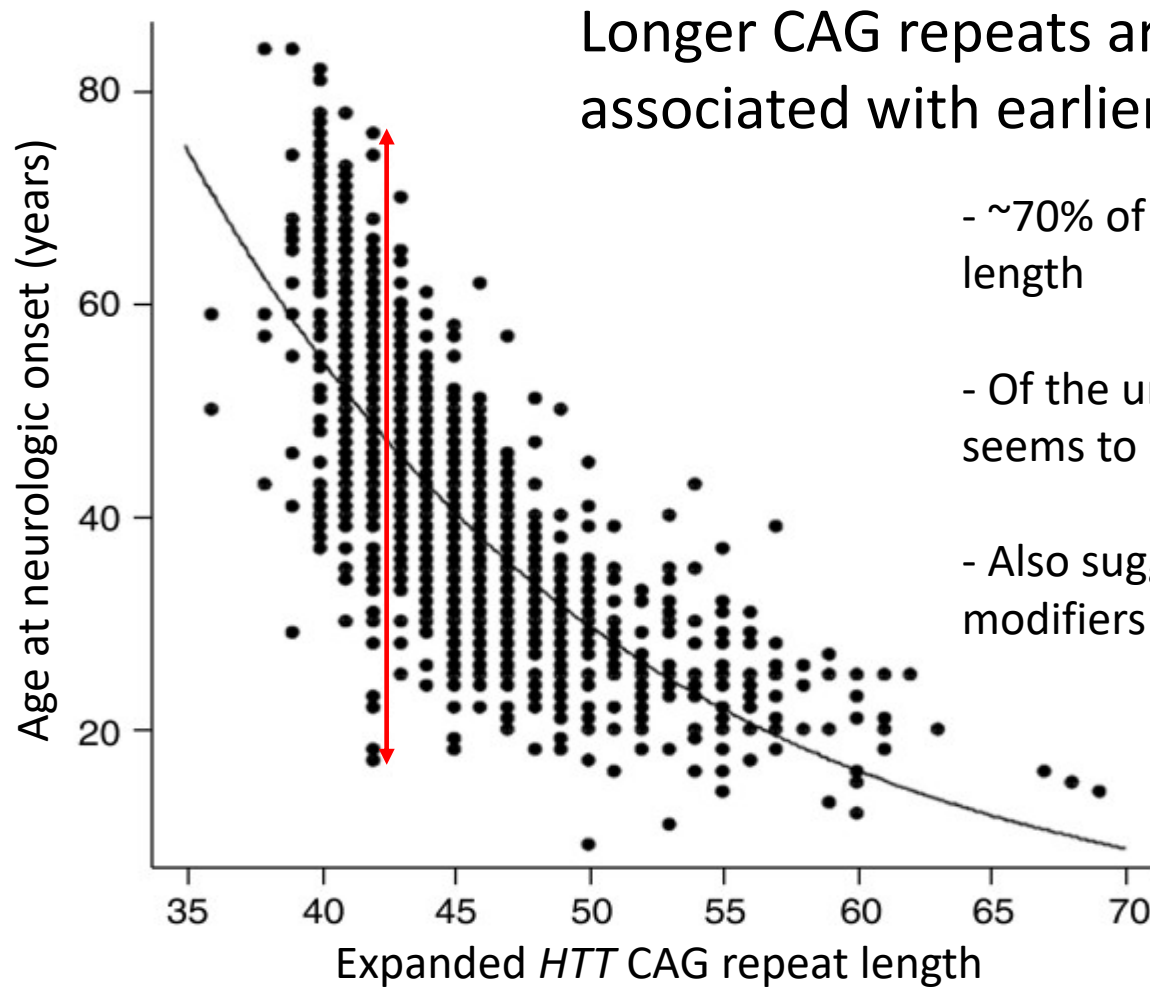


- HD is characterized by progressive neuronal cell death associated with mood swings, choreic movements and progressive dementia
- Anticipation: earlier age of onset from one generation to the next



1. The Huntington's Disease Collaborative Research Group (1993) *Cell*, 72 (6): 971-83
2. Harvard Brain Tissue Resource Center

# Inverse correlation of age at neurologic onset and mutant CAG repeat length



- ~70% of variation explained by CAG length
- Of the unexplained variation, 40% seems to be heritable.
- Also suggesting the case for genetic modifiers of HD age of onset.

# HD Age at Onset GWAS

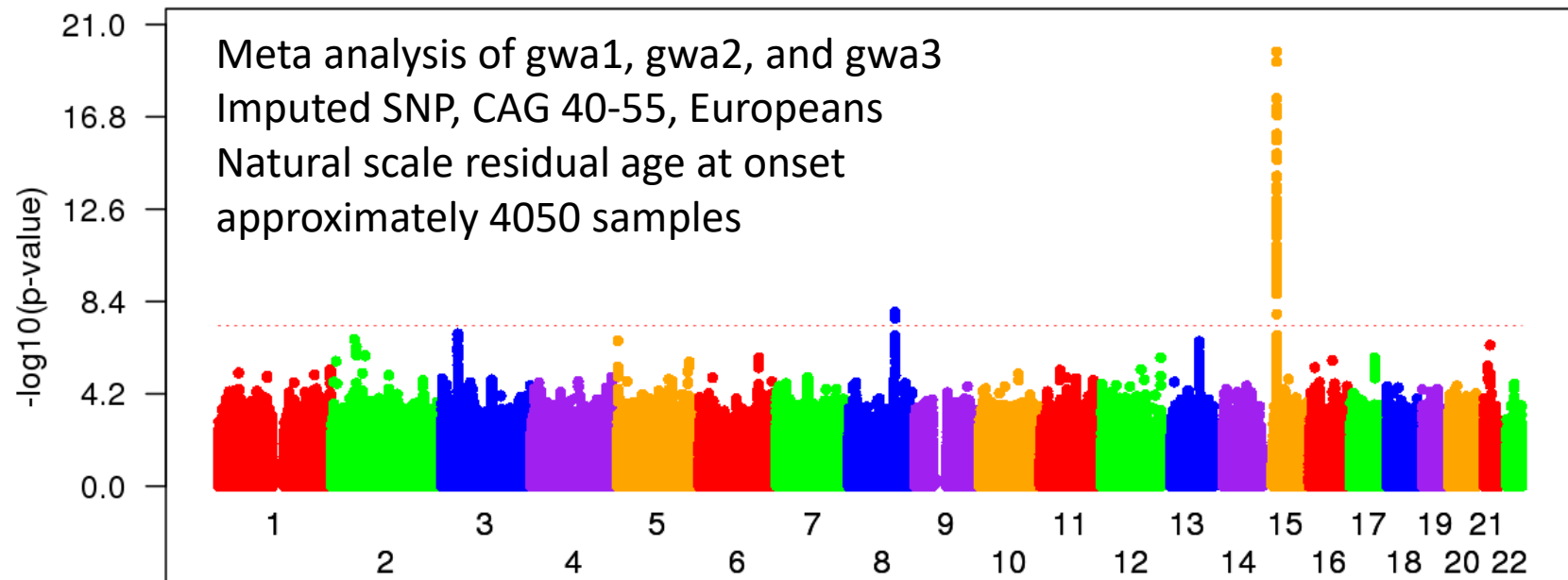
Cell

## Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease

Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium\*

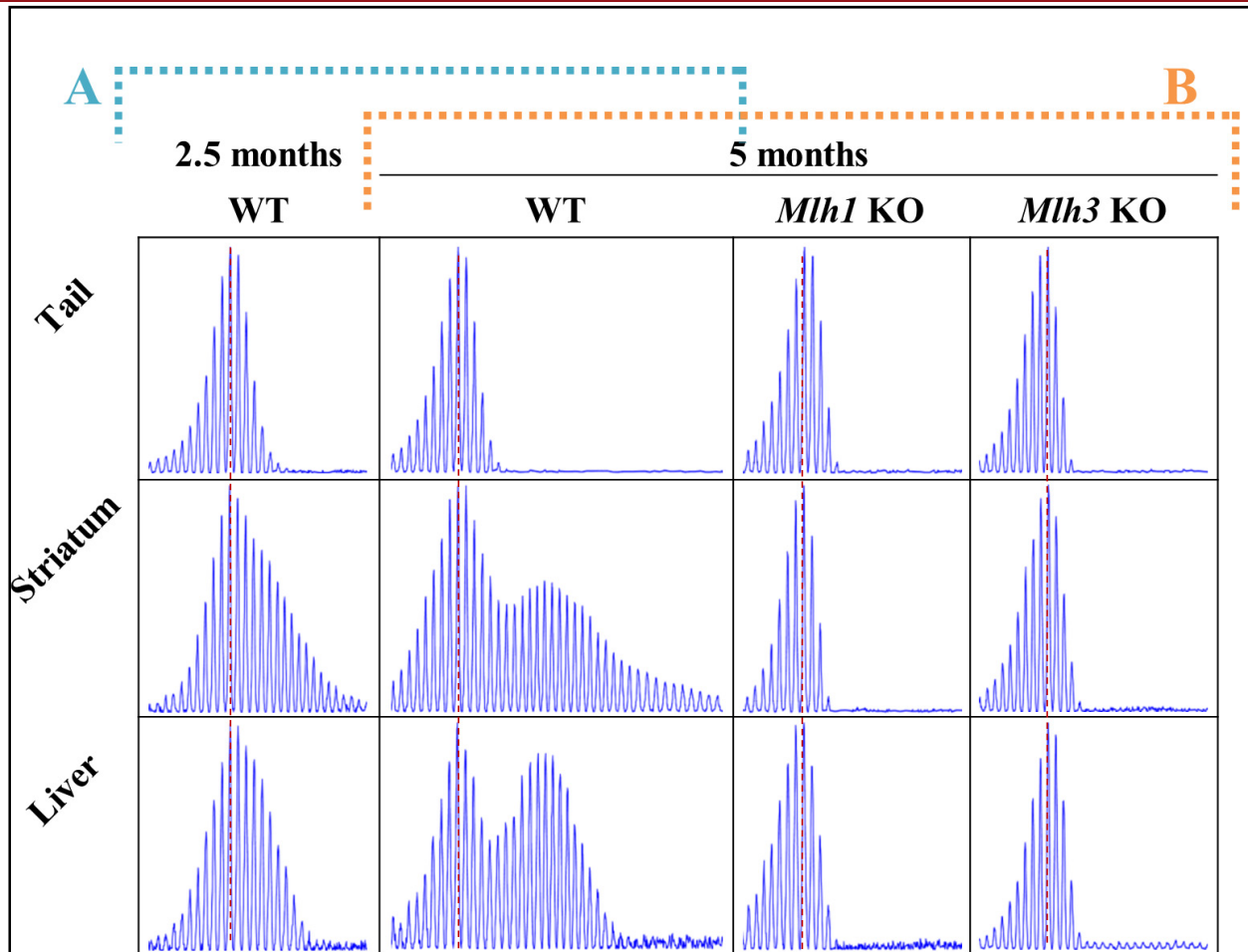
\*Correspondence: [gusella@helix.mgh.harvard.edu](mailto:gusella@helix.mgh.harvard.edu)

<http://dx.doi.org/10.1016/j.cell.2015.07.003>



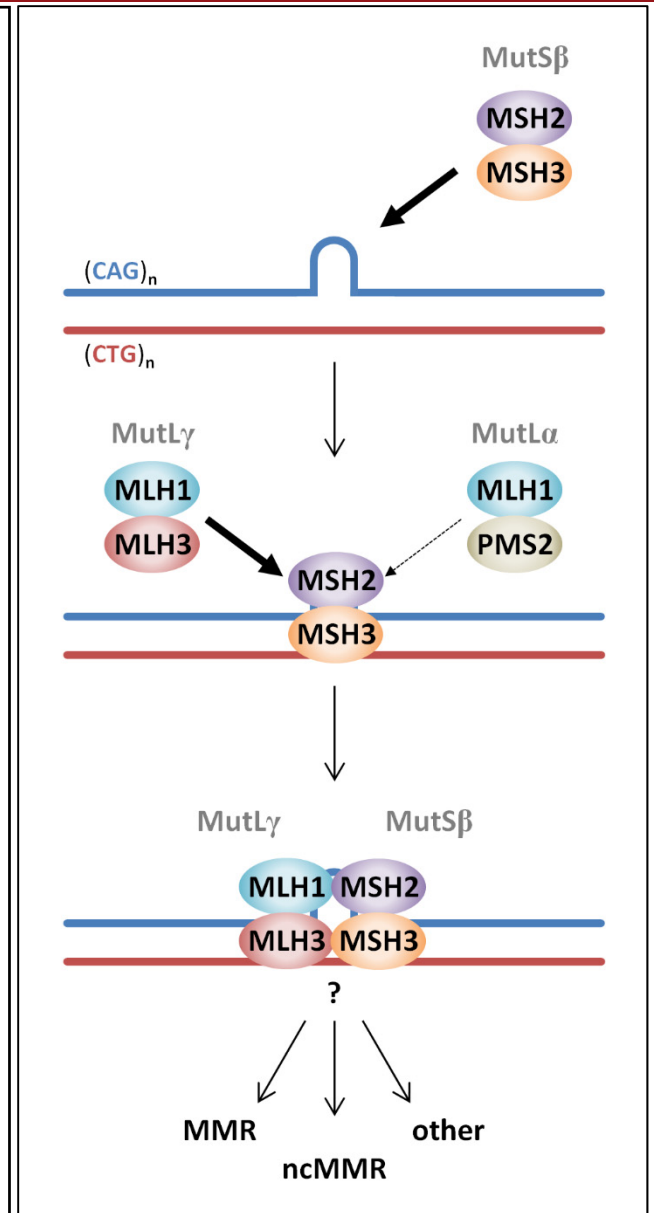
- *MLH1* variants with nominal significance
- Pathway analysis implicates multiple DNA repair/handling genes

# Mismatch repair and CAG instability

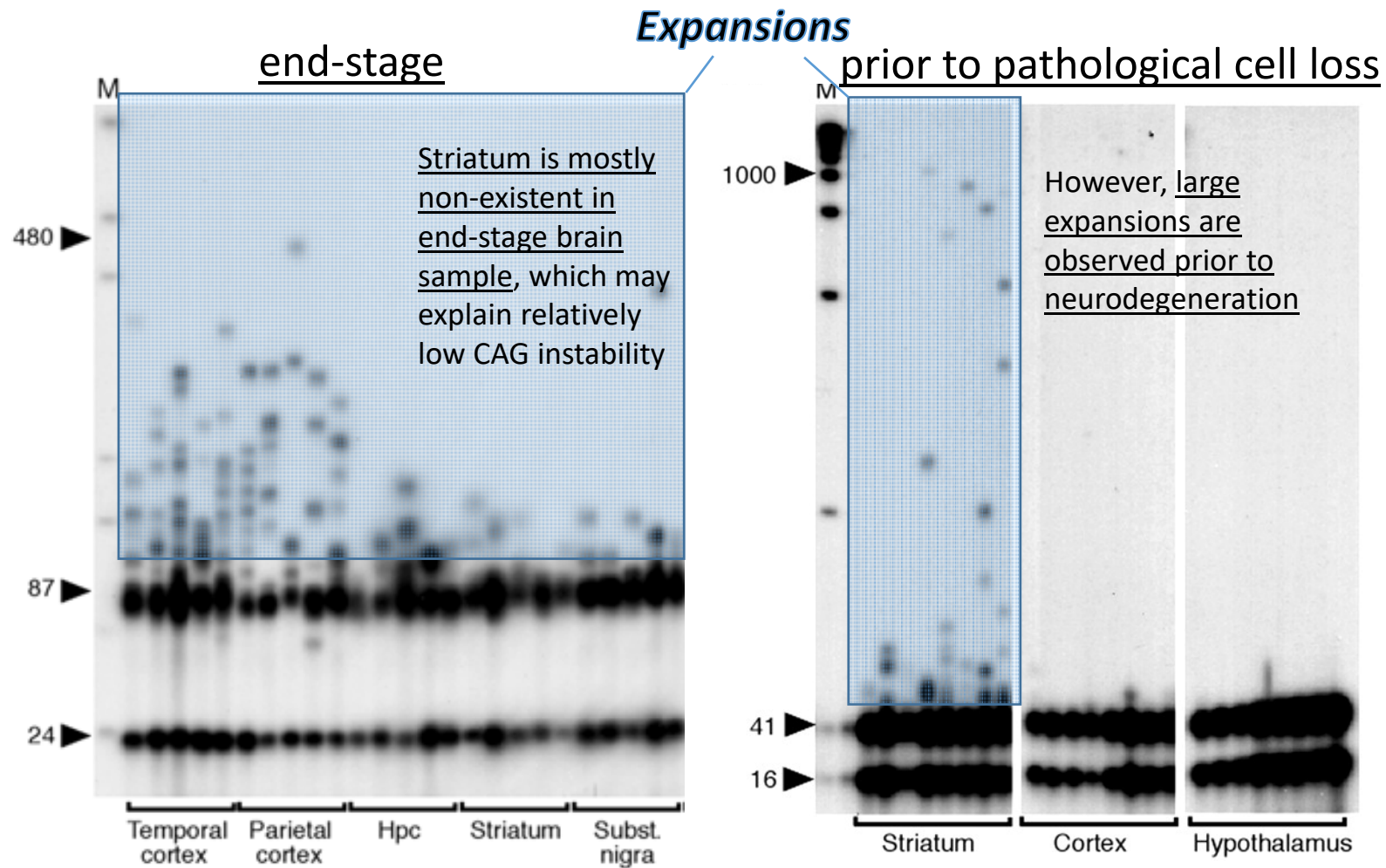


**A)** Age-dependent and tissue-specific CAG instability in the striatum and liver of  $Hdh^{Q111}$  mice;

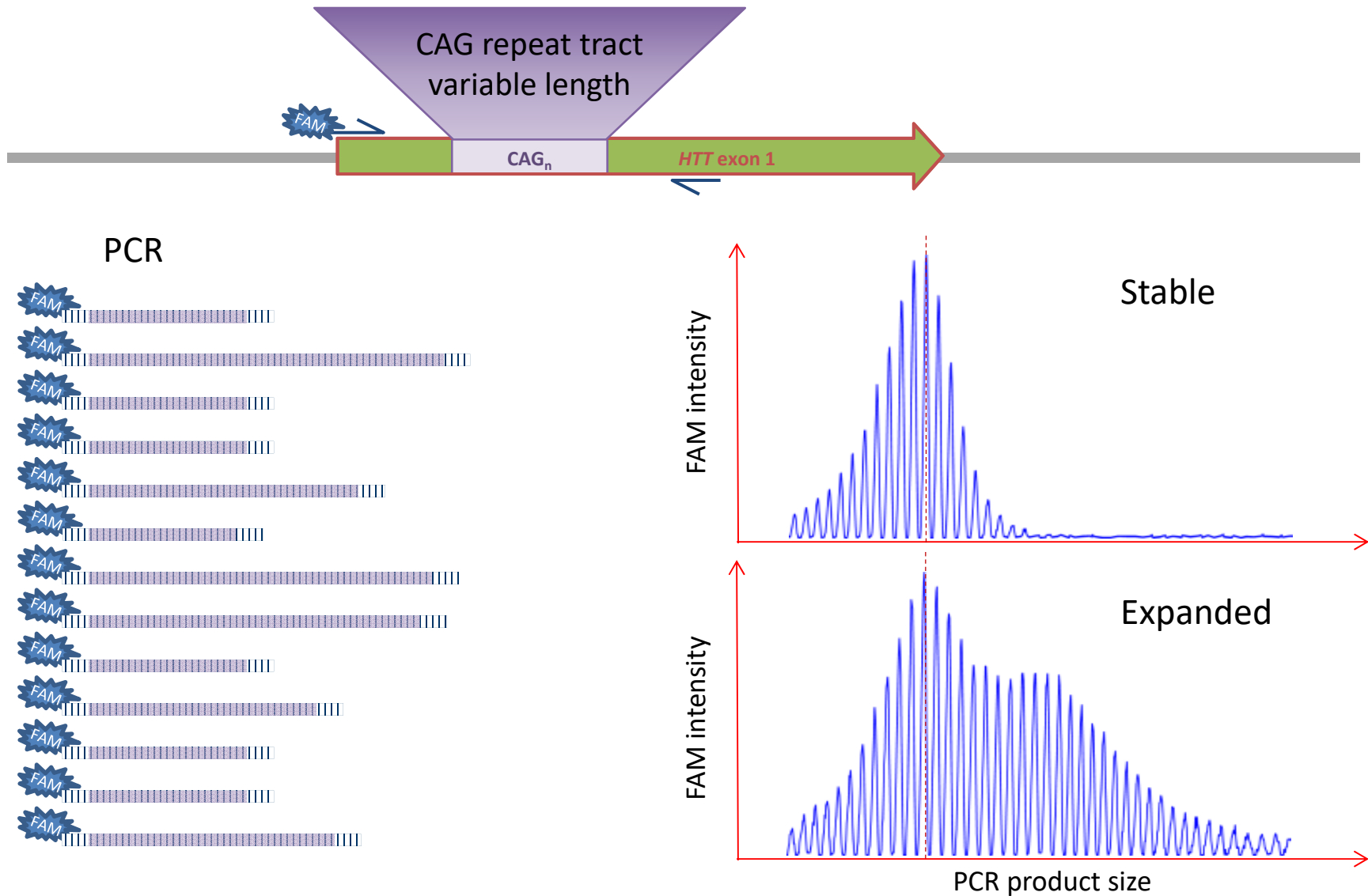
**B)** *Mlh1* and *Mlh3* are enhancers of somatic CAG expansion.



# Somatic CAG instability in HD patient postmortem brain



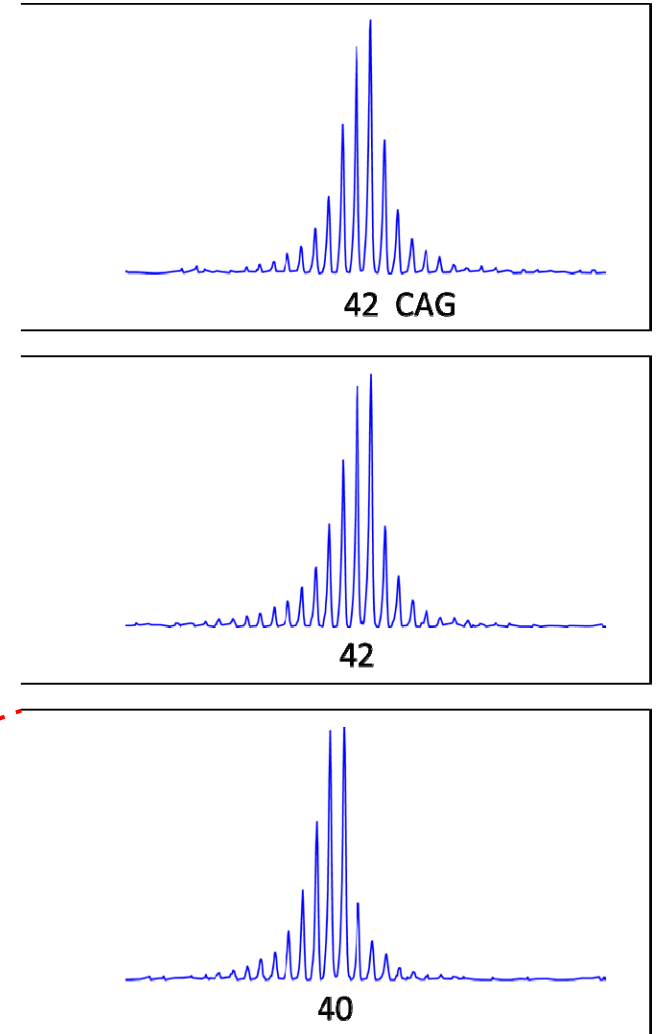
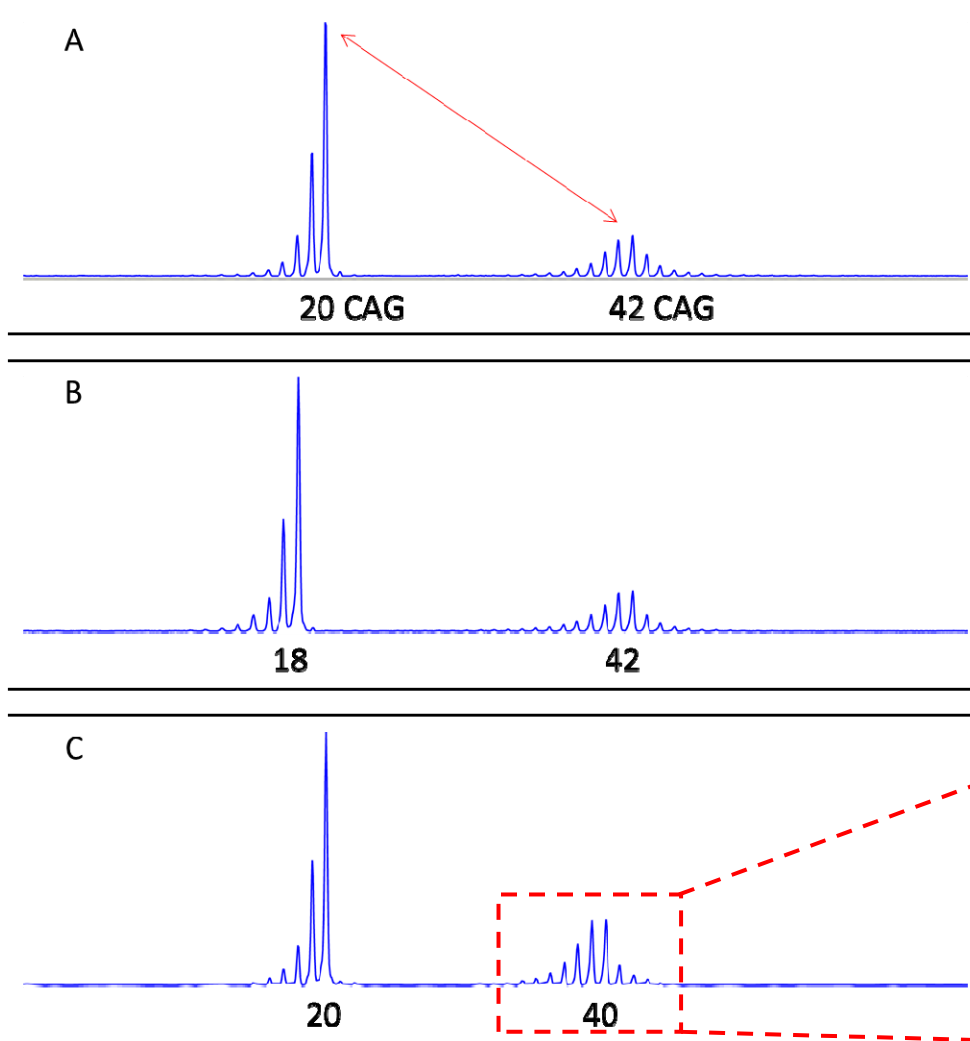
# GeneScan analysis



# GeneScan analysis

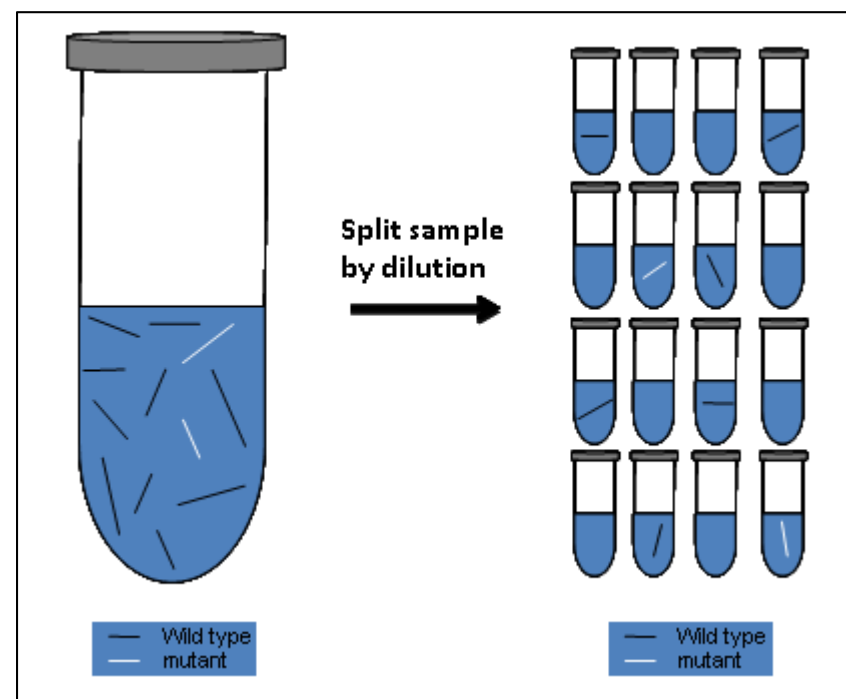
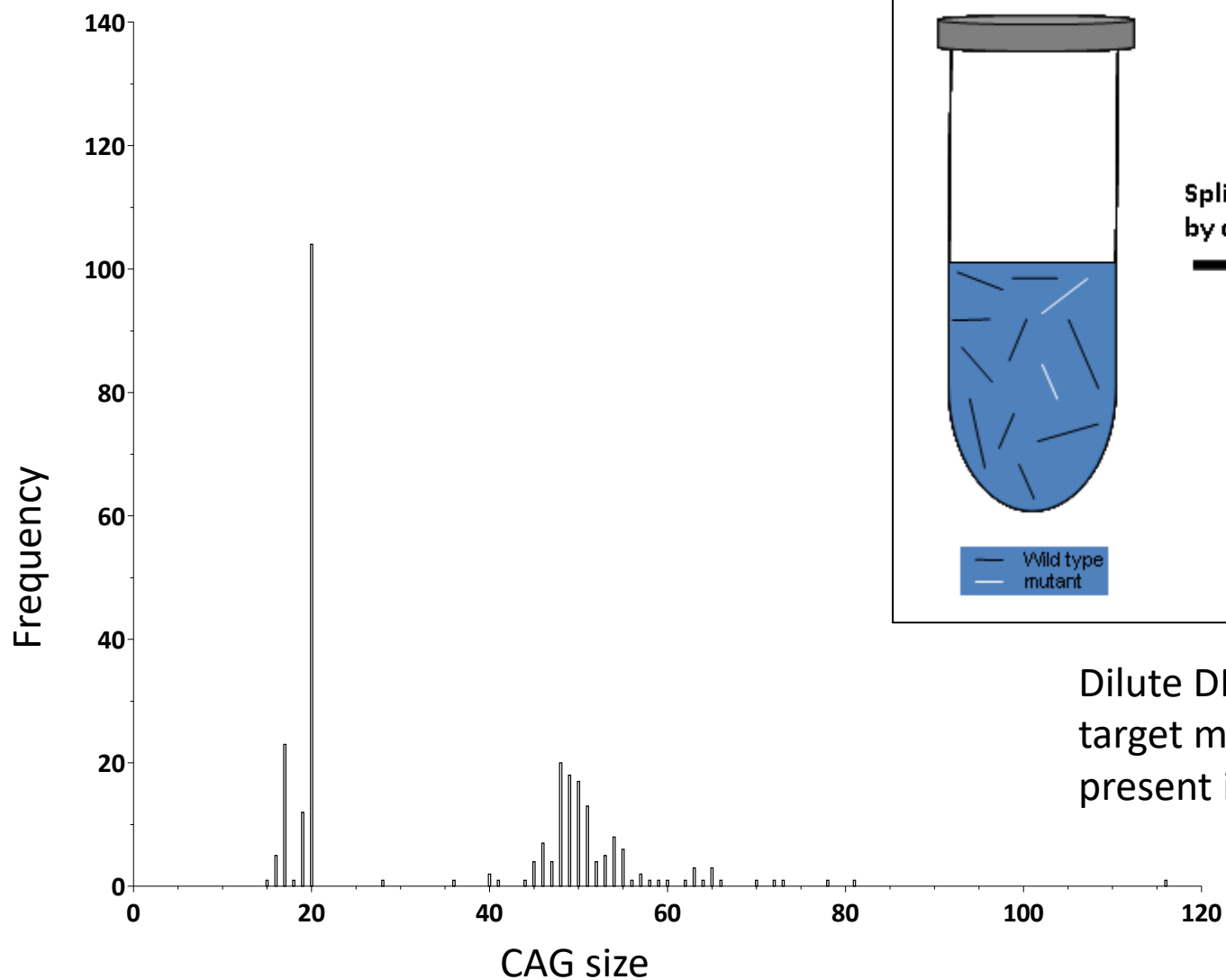
Strong bias towards PCR amplification of smaller (Wt) allele

Lack of sensitivity capturing low frequency expanded alleles



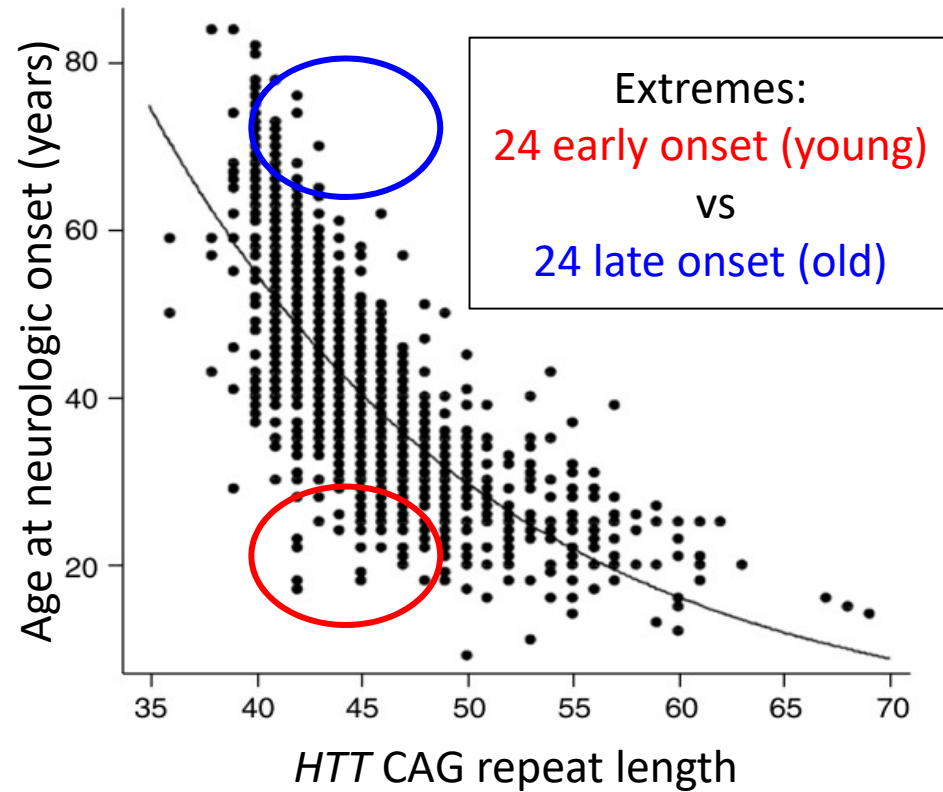


# Small-pool PCR

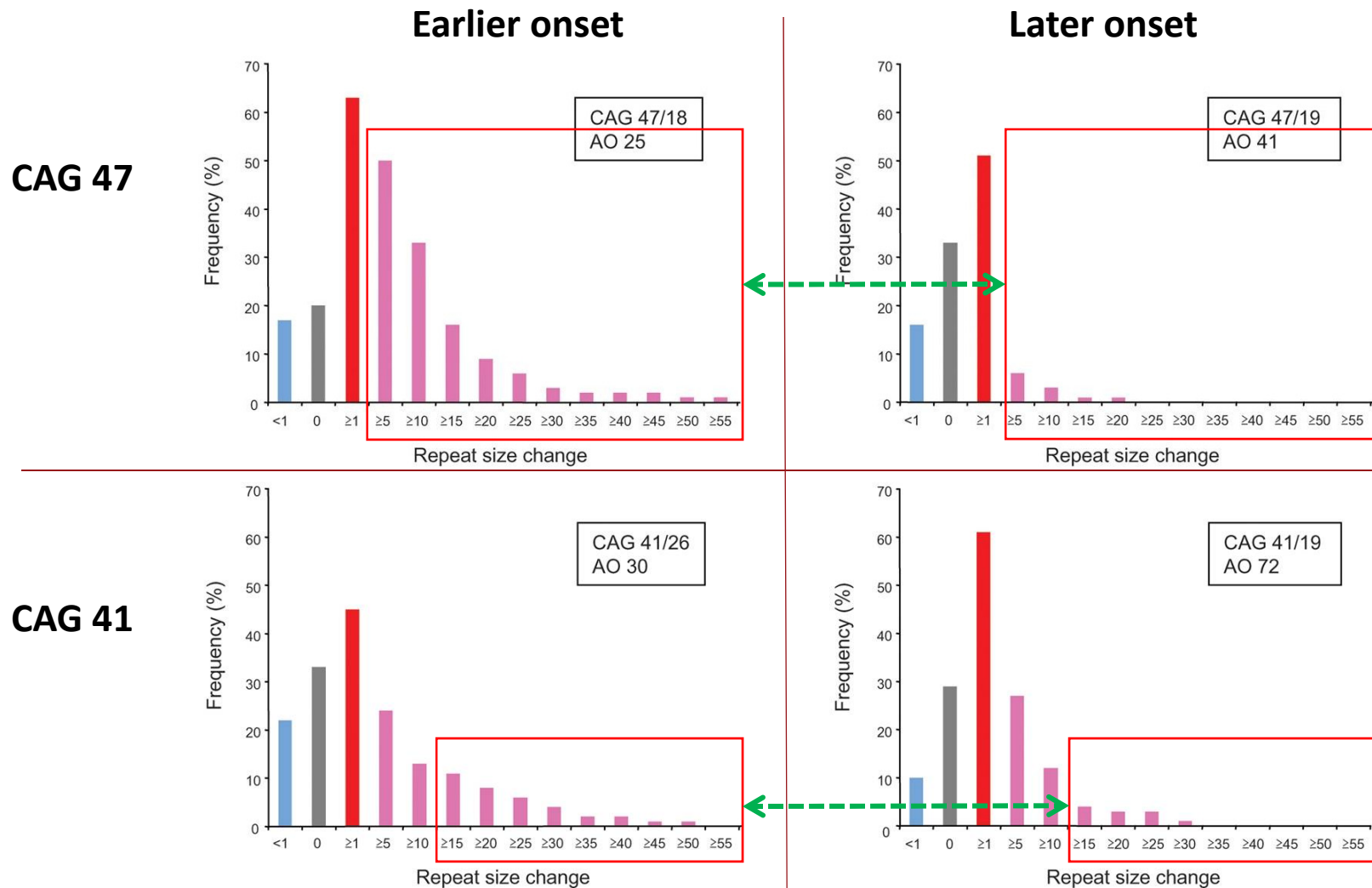


Dilute DNA so that very few target molecules are present in PCR reaction.

# CAG repeat length vs HD age of onset



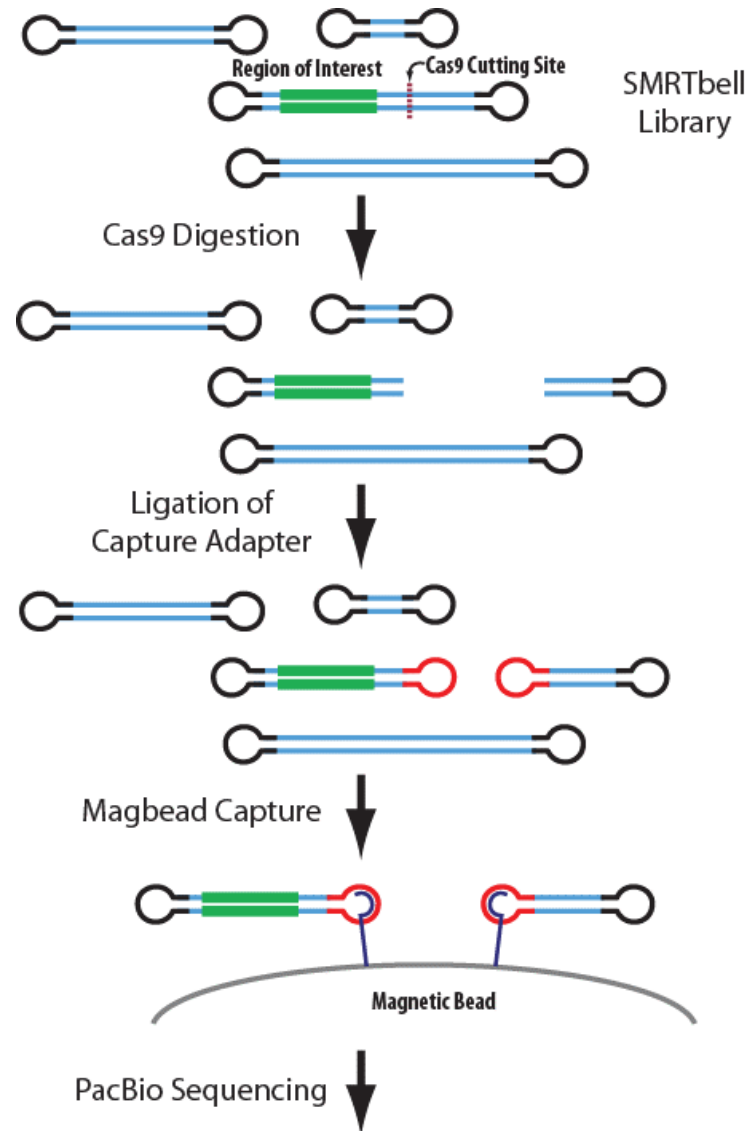
# Somatic CAG expansions in the brain are associated with earlier age of HD onset



## ***MLH1* variant effect on HD onset and somatic CAG expansions**

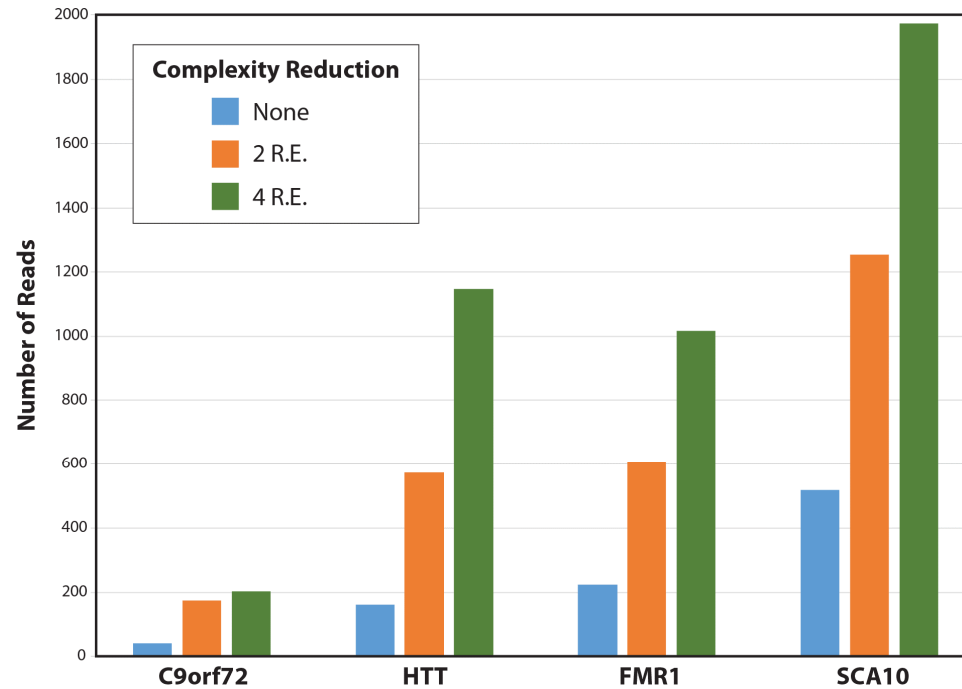
- **Suggestive effect of *MLH1* rare variant but not enough samples**
- **SP-PCR too laborious and not cost effective**

# Amplification free target enrichment



# Amplification free target enrichment

Complexity reduction increases the number of on-target reads

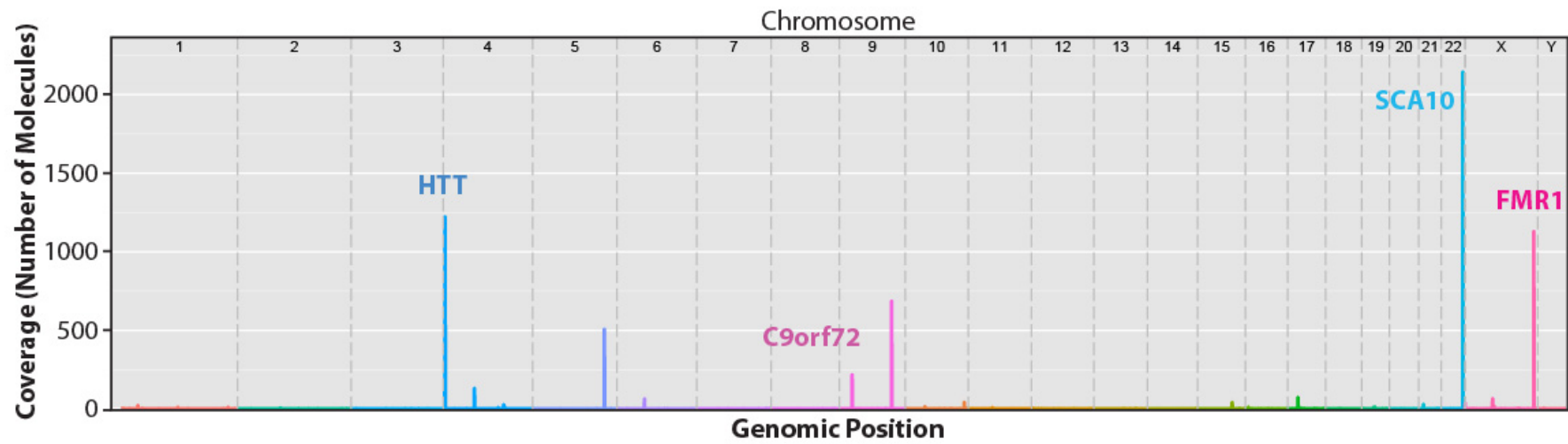


Complexity Reduction Library and Sequencing Yield

Complexity Reduction	Input Genomic DNA	Final SMRTbell Yield	% Yield	CCS Reads	On-Target Reads	% Reads On-Target
None	18.0 µg	6.7 µg	37.2%	44,031	945	2.15%
2 R.E.	20.0 µg	3.0 µg	15.0%	51,806	2609	5.04%
4 R.E.	20.0 µg	1.6 µg	8.0%	45,676	4335	9.49%

# Amplification free target enrichment

## Coverage across genome



# Amplification free target enrichment

- BamHI-EcoRI SB library prep and Cas9 enrichment experiment:

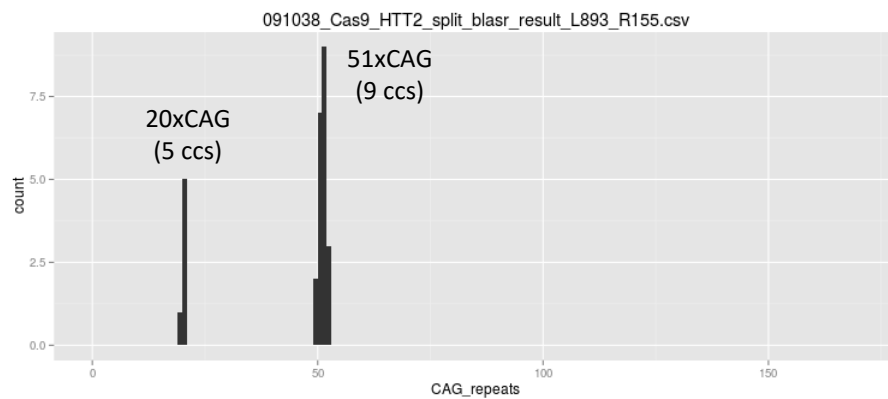
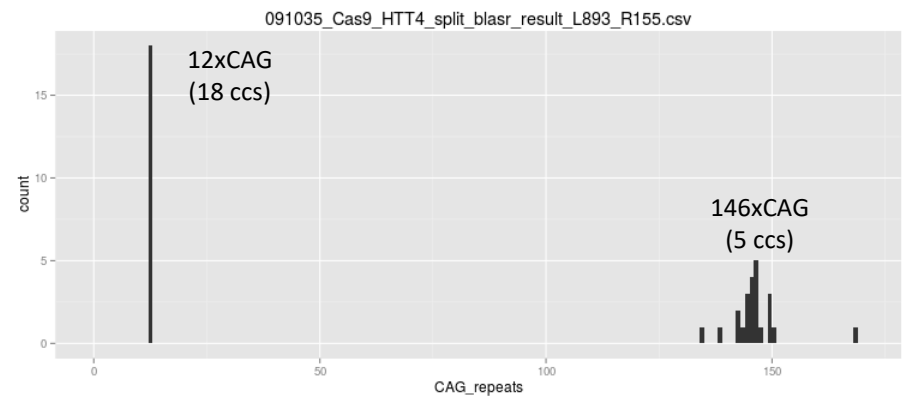
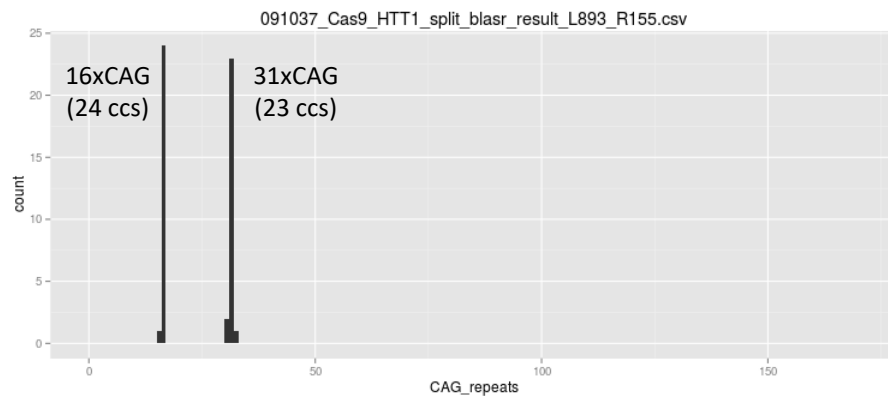
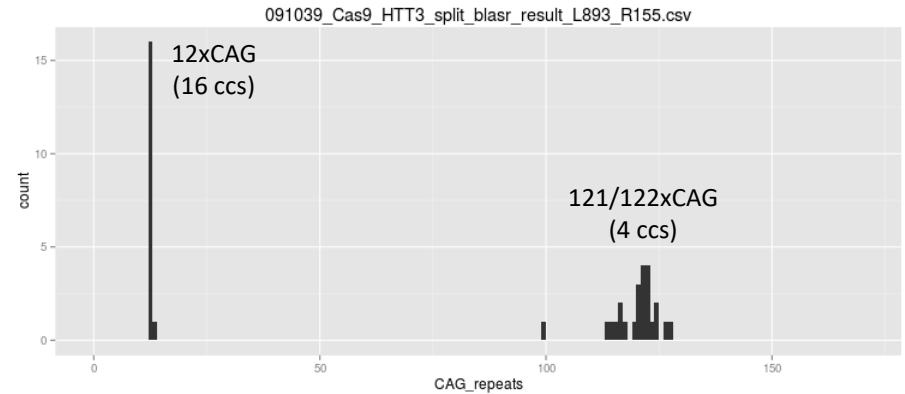
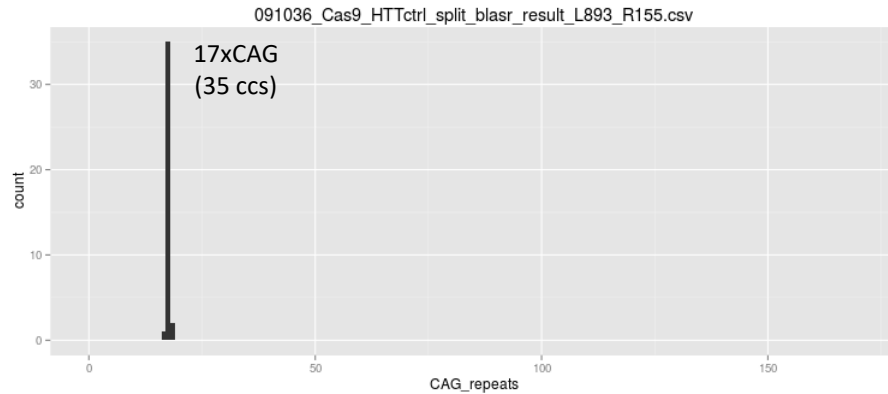
	HTT control	HTT1	HTT2	HTT3	HTT4
Input gDNA	-	4 µg	4 µg	4 µg	4 µg
BamHI-EcoRI SB	Pre-made SB	2.4 µg	0.9 µg	1.8 µg	0.9 µg
Input for Cas9 exp.	1 µg	1 µg	0.9 µg	1 µg	0.9 µg

- 1 SMRTcell sequencing with P6C4 and 6hr movies

Sample Name	SMRT Cells	Total Reads	Polymerase Reads			Reads Of Insert		
			Length	Quality	Mbases	Length	Quality	Mbases
Cas9_HTT_ctrl_P6C4_6hr	1	17119	10228	0.83	175.1	2936	0.87	50.3
Cas9_HTT1_P6C4_6hr	1	21136	11888	0.83	251.3	3233	0.88	68.3
Cas9_HTT2_P6C4_6hr	1	26823	13633	0.83	365.7	3321	0.89	89.1
Cas9_HTT3_P6C4_6hr	1	28935	13325	0.83	385.6	3403	0.89	98.5
Cas9_HTT4_P6C4_6hr	1	24327	13366	0.83	325.2	3187	0.89	77.5



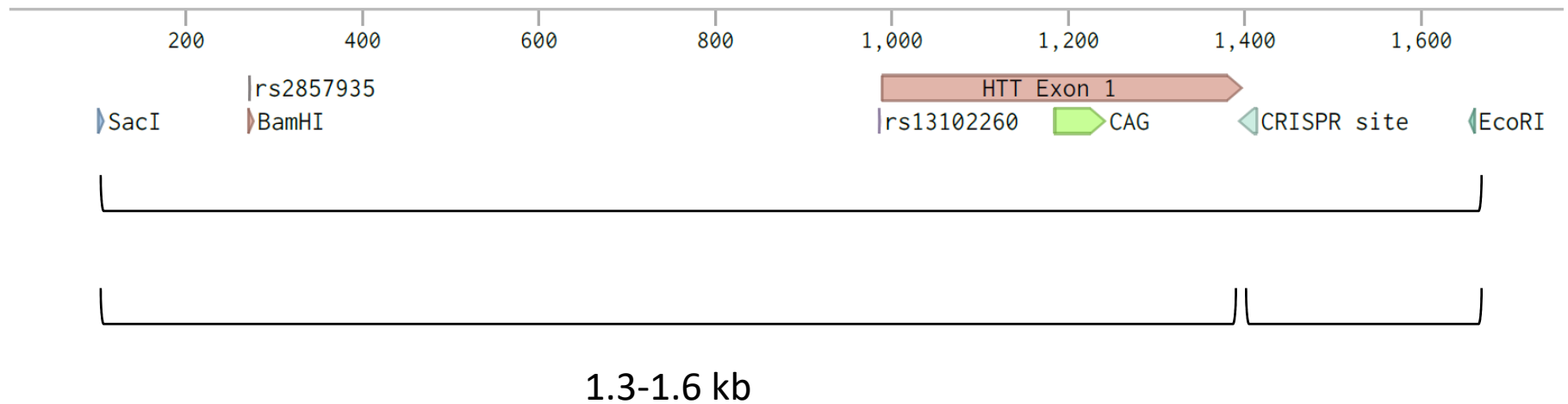
# Amplification free target enrichment



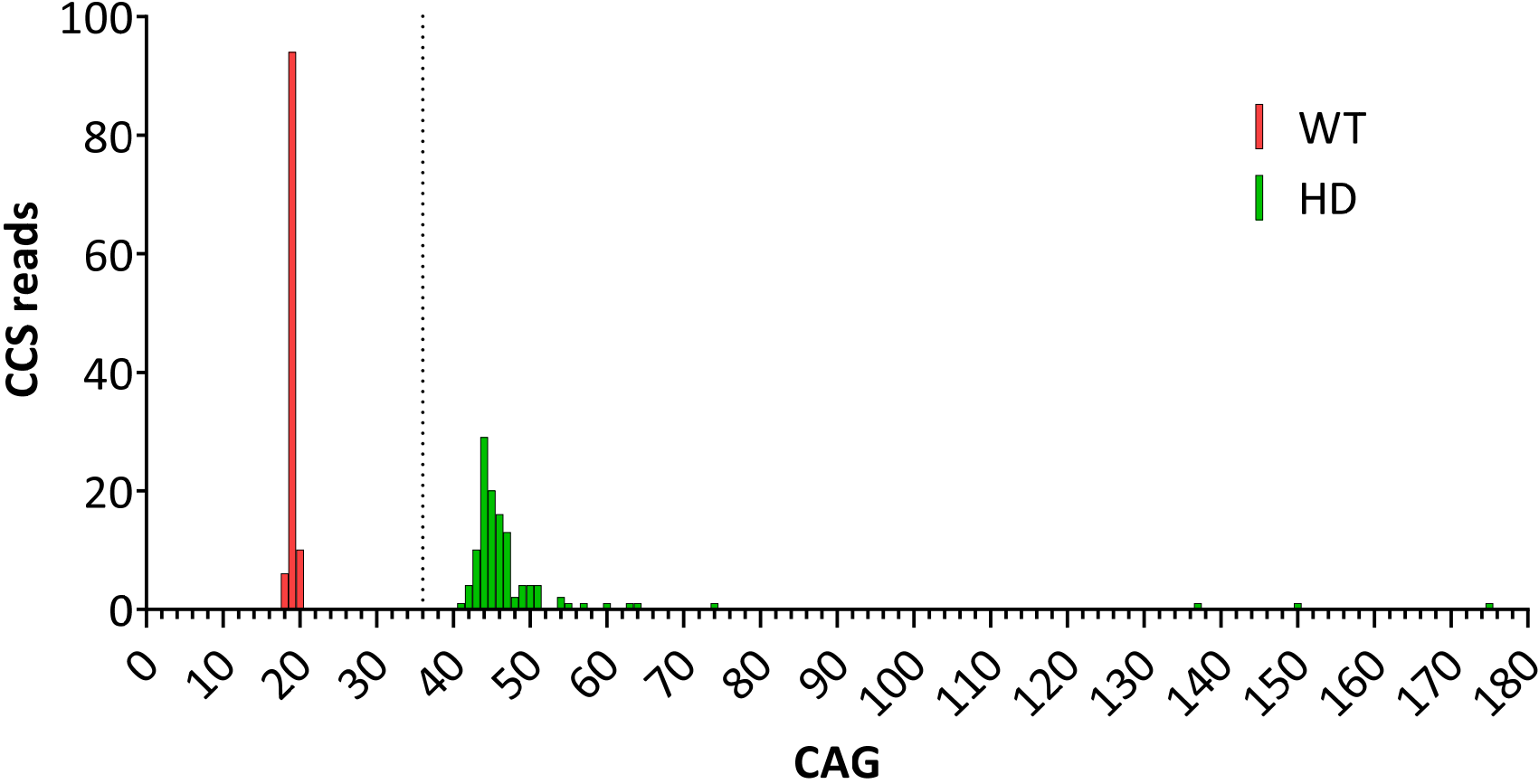


# Amplification free target enrichment

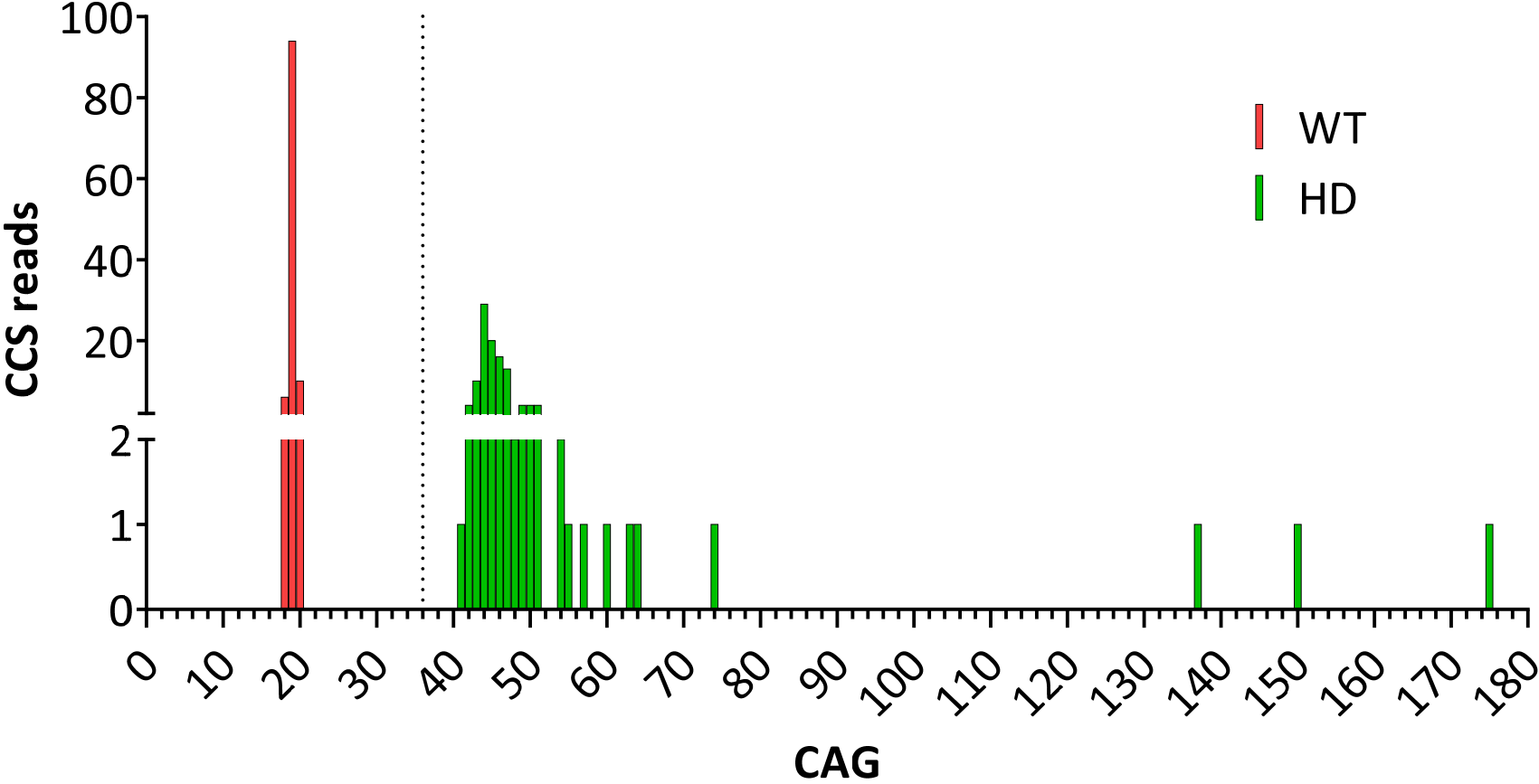
## *HTT* target locus



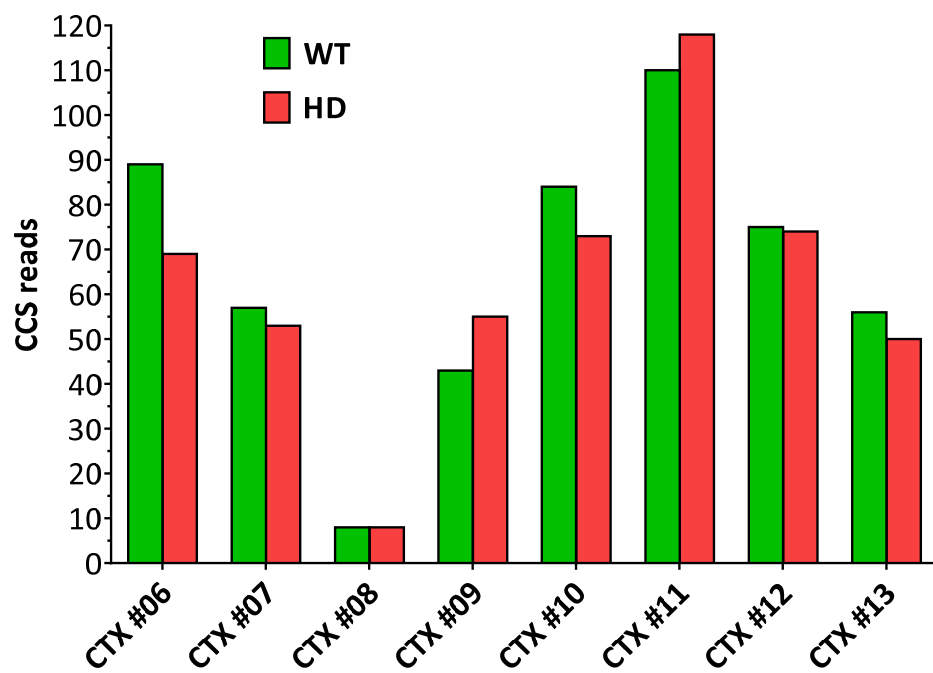
# CAG repeat sizing in HD patient post-mortem brain



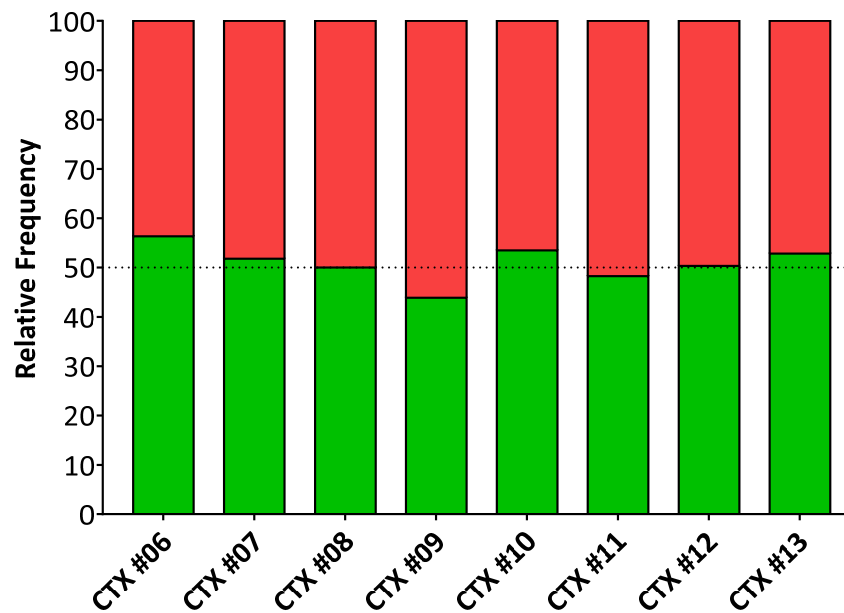
# CAG repeat sizing in HD patient LCLs



# CAG repeat sizing in HD patient LCLs

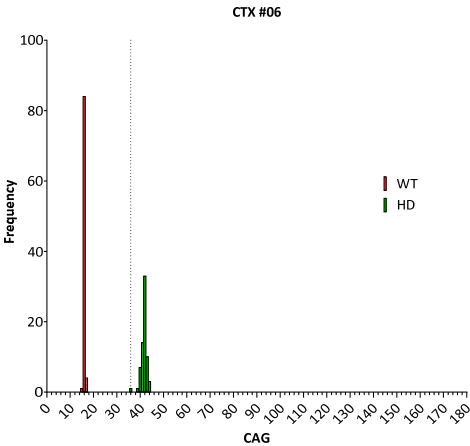


No bias observed in terms of capturing WT vs HD alleles

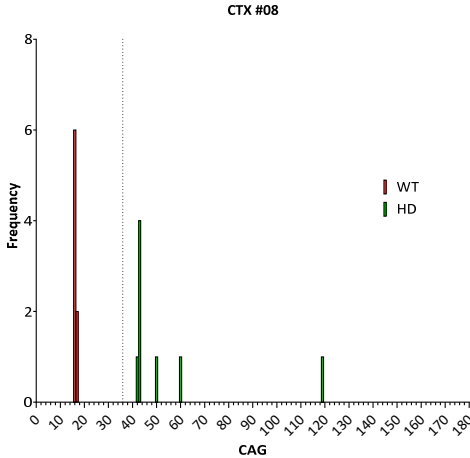


# CAG repeat sizing in HD patient LCLs

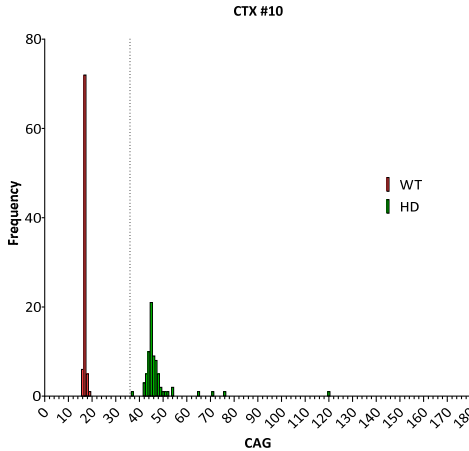
## CAG 42



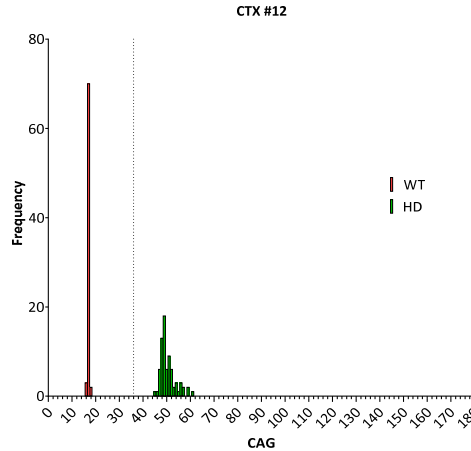
## CAG 43



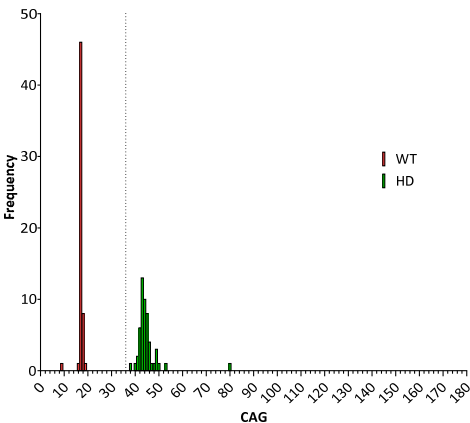
## CAG 44



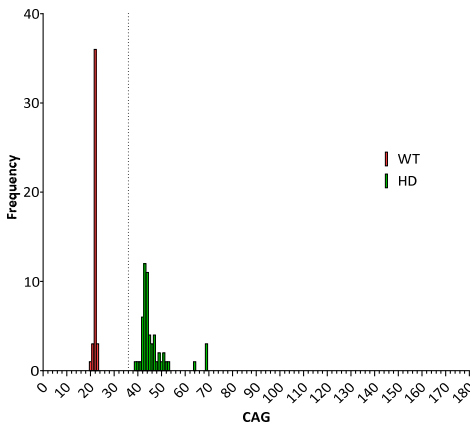
## CAG 47



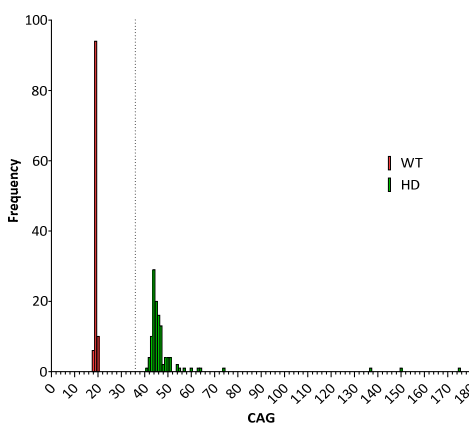
## CAG 42



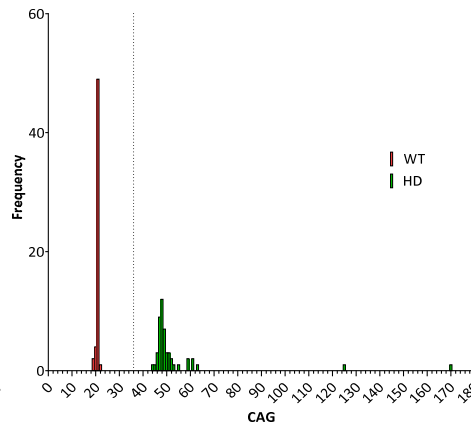
## CAG 43



## CAG 44

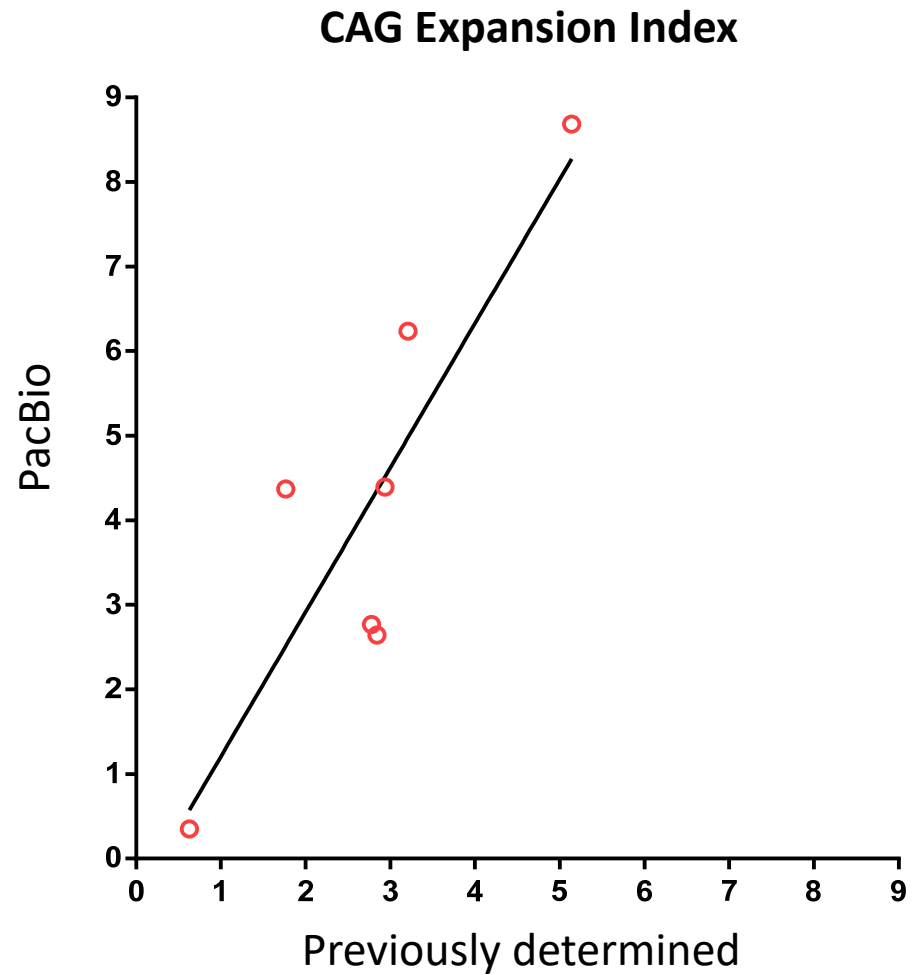


## CAG 47



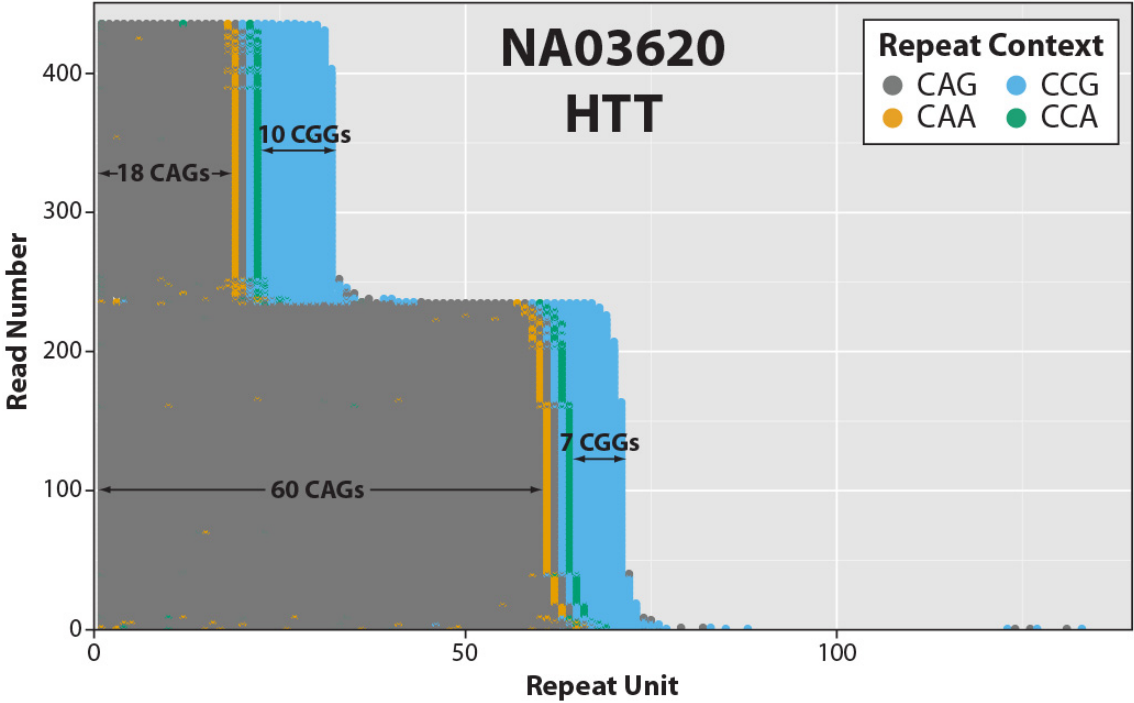
# CAG repeat sizing in HD patient LCLs

Strong correlation with previous data





# Single molecule resolution



Interruptions and polymorphic proline repeat

# Amplification free target enrichment

- **Disadvantages**

- Need significant amount of gDNA (>5ug)
- More susceptible to gDNA quality
  - postmortem brain: variable time to collection

- **Advantages**

- No apparent length-related bias
- Ability to detect large CAG alleles
- Sequence-level resolution
  - repeat interruptions?
  - *cis*-modifiers: phased alleles
    - polymorphic proline repeat and other variants

# Acknowledgements

- **CGM:**

- Vanessa Wheeler
- James Giordano
- **Jong-Min Lee**
- Marcy MacDonald
- Jim Gusella
- CGM Genotyping Core

- **PacBio:**

- Tyson Clark
- **Yu-Chih Tsai**
- Jonas Korlach
- Igor Vilfan
- Thang Pham
- Michael Weiland



Harvard **NeuroDiscovery** Center

COLLABORATING TO CURE  
NEUROLOGIC AND PSYCHIATRIC DISEASE

