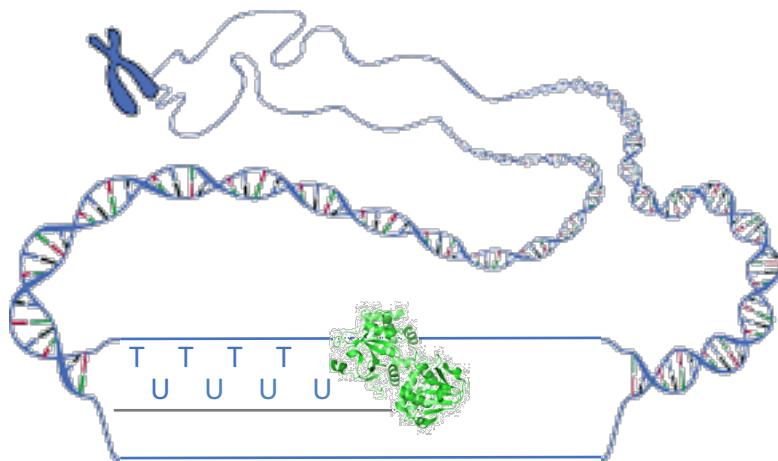


# Short Tandem Repeats not just junk any more!

Tugce BILGIN SONAY

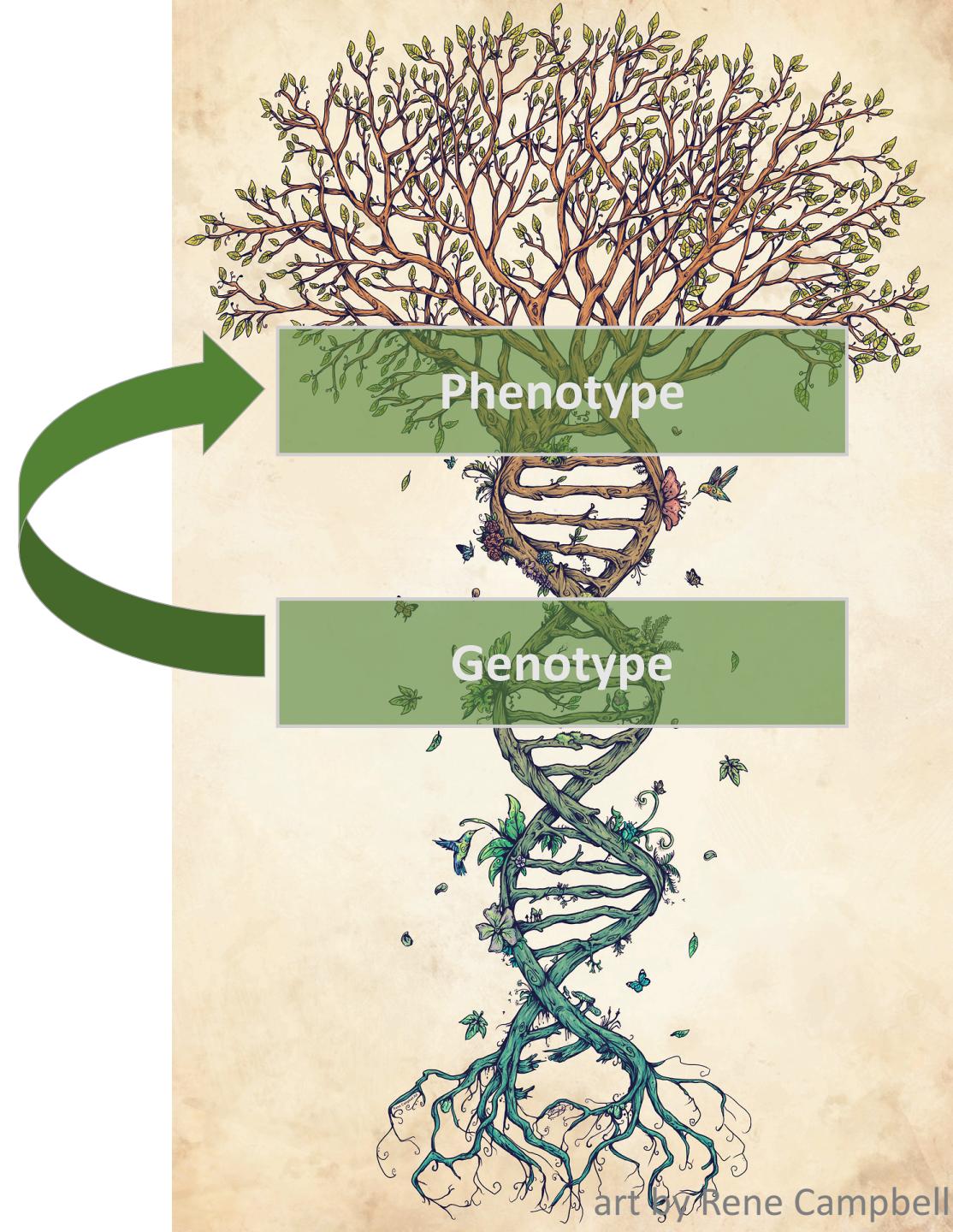


## Genetic Determinants of Phenotypic Divergence:

- **Single Nucleotide Polymorphisms**

1 mutation in every 10-100 million nucleotides

- **Structural Variants** – large deletions, insertions 5% of human genome



## Genetic Determinants of Phenotypic Divergence:

- Single Nucleotide Polymorphisms

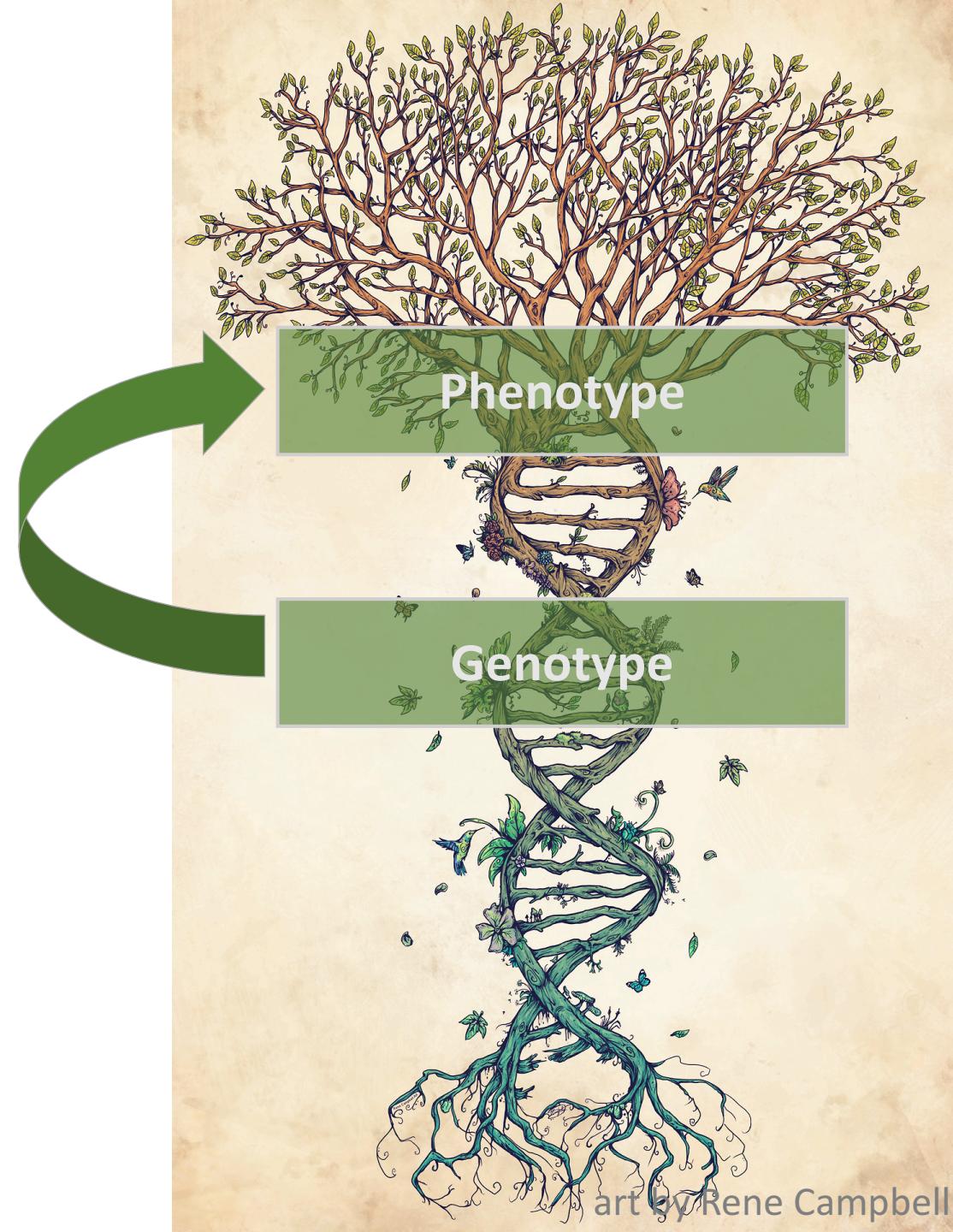
1 mutation in every 10-100 million nucleotides

- Structural Variants – large deletions, insertions 5% of human genome

- Tandem repeat instability

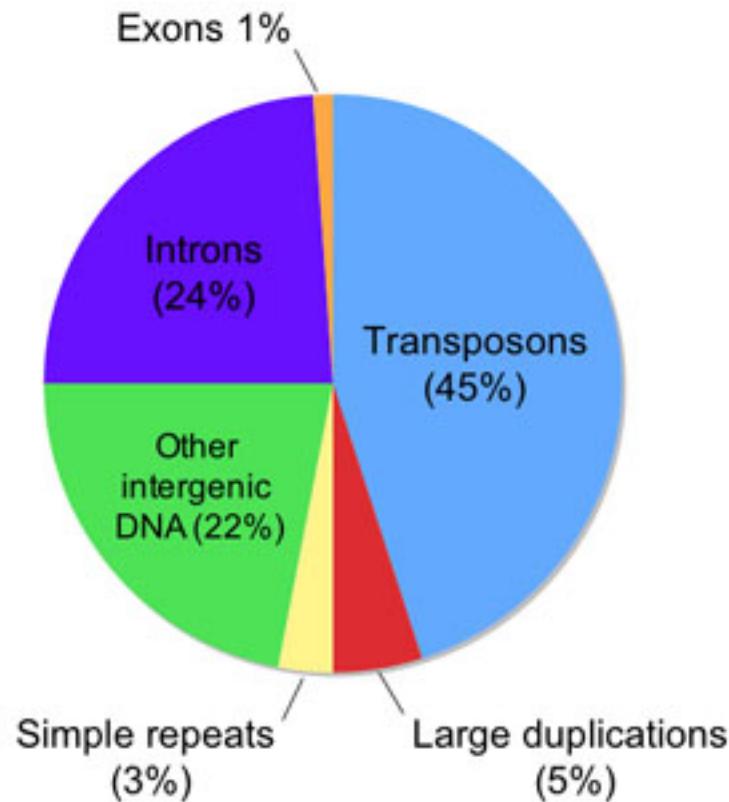
20 % of functional variants

...ATTTAATTTAATTTAATTTA...



# What is an STR?

**Short tandem repeat,  
also known as  
microsatellite**

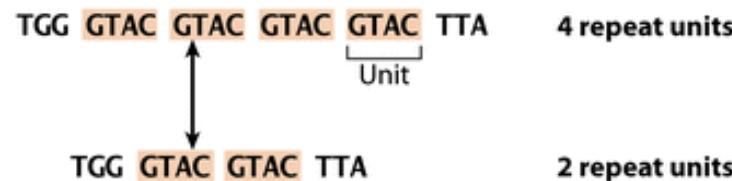


# What is an STR?

Tandem repeats characteristics

- Unit size
- Number of units

Example: 1 unit = GTAC (tetranucleotide)



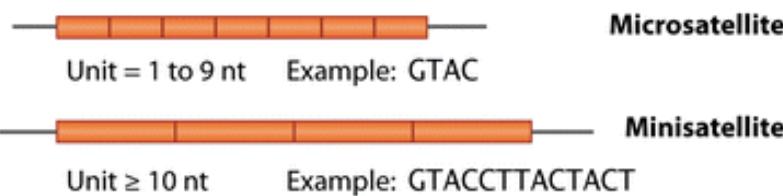
## Purity

100% GTAC GTAC GTAC GTAC GTAC GTAC

96% GTAC GTAC GTAC GTAC GAAC GTAC

79% GTAC CTAC GTGC GTAC GAAC CAAC

## Unit length



Gemayel R, et al. 2010.

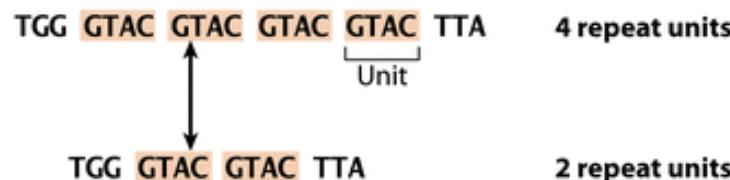
Annu. Rev. Genet. 44:445–77

# What is an STR?

Tandem repeats characteristics

- Unit size
- Number of units

Example: 1 unit = GTAC (tetranucleotide)



## Purity

100% GTAC GTAC GTAC GTAC GTAC GTAC

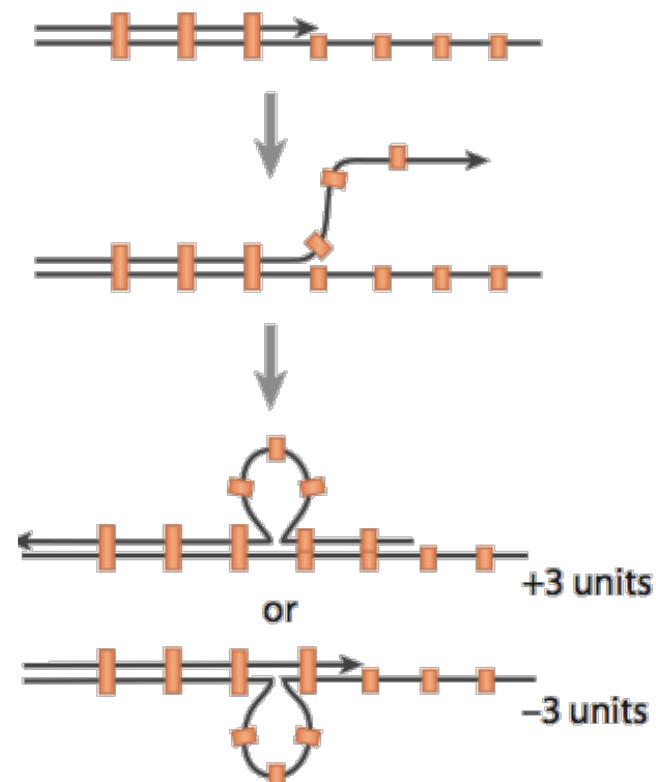
96% GTAC GTAC GTAC GTAC GAAC GTAC

79% GTAC CTAC GTGC GTAC GAAC CAAC

## Unit length



extremely high mutation rates:  $10^{-2}$ - $10^{-5}$  per cell division



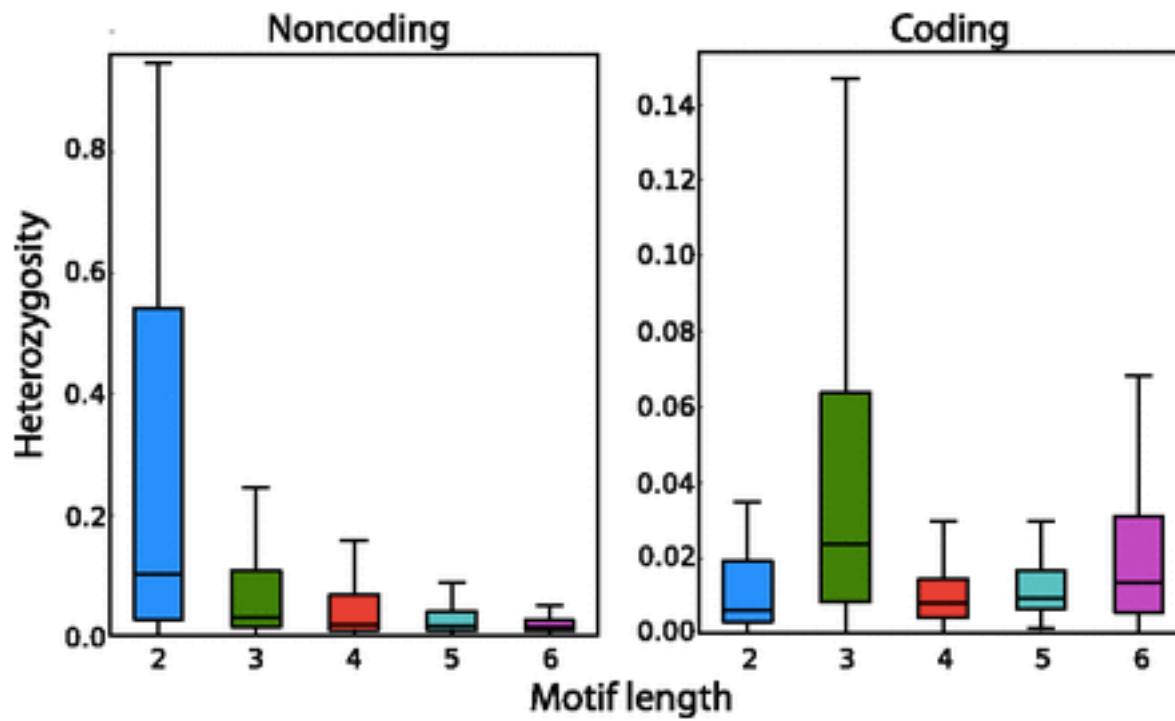
Replication  
slippage

# STR variation

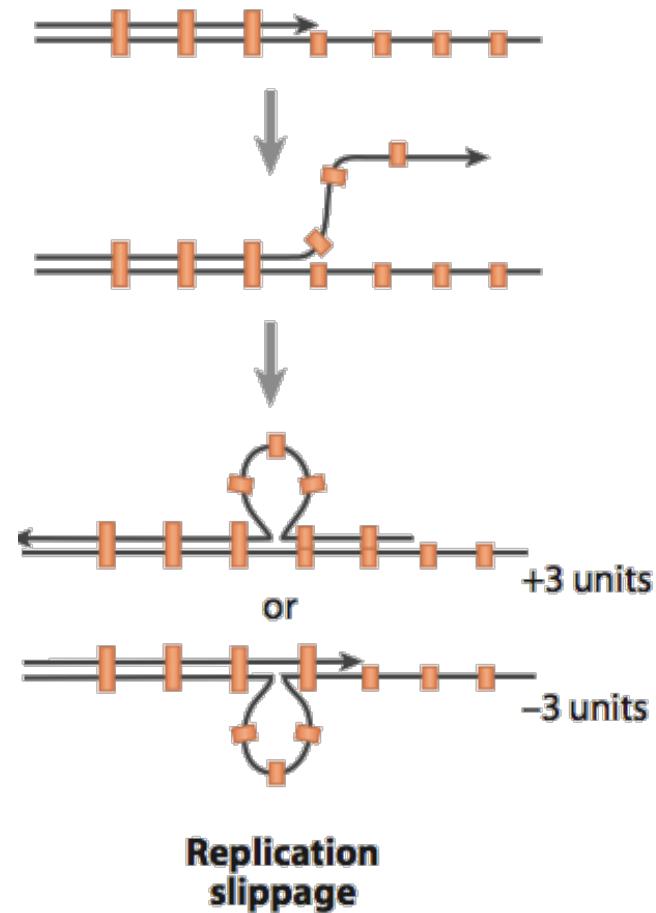
Resource

## The landscape of human STR variation

Thomas Willems,<sup>1,2</sup> Melissa Gymrek,<sup>1,3,4,5</sup> Gareth Highnam,<sup>6</sup>  
The 1000 Genomes Project Consortium, David Mittelman,<sup>6,7</sup> and Yaniv Erlich<sup>1</sup>



extremely high mutation rates:  $10^{-2}$ - $10^{-5}$  per cell division



# masking, masking, masking - ignorance is a bliss

## REPEAT MASKER

<http://www.repeatmasker.org>

1. Copy and Paste Sequence

2. DNA Source  
(Arabidopsis thaliana)

3. Submit Sequence

Systems Biology RepeatMasker Web Server

RepeatMasker screens DNA sequences in FASTA format against a library of repetitive elements and returns a masked query sequence ready for database searches. RepeatMasker also generates a table annotating the masked regions.

Reference: A.P.A. Smit, R. Hubley & P. Green, unpublished data. Current Version: open-3.2.7 (RMLib: 20090120)

[Check Current Queue Status](#)

**Basic Options**

Sequence:  Choose File no file selected  
OR  
Select a sequence file to process or paste the sequences(s) in FASTA format. Large sequences will be queued, and may take a while to process.

Search Engine:  wblast  cross\_match

Speed/Sensitivity:  rush  quick  default  slow

DNA source: Human

Return Format:  html  tar file

Return Method:  html  email Your email address

## Pre-processing: repeat masking

### Repeated elements:

- They represent a big part of the mammalian genome.
- They are found in a number of genomes (plants, ...)
- They induce errors in clustering and assembling.
- They should be masked, not deleted, to avoid false sequence assembling.
- ... but also interesting elements for evolutionary studies.
- SSRs important for mapping of diseases.

### Tools to find repeats:

- RepeatMasker has been developed to find repetitive elements and low-complexity sequences. RepeatMasker uses the cross-match program for the pairwise alignments
  - <http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker>
- MaskerAid improves the speed of RepeatMasker by ~ 30 folds using WU-BLAST instead of cross-match
  - <http://sapiens.wustl.edu/maskeraid>
- RepBase is a database of prototypic sequences representing repetitive DNA from different eukaryotic species:
  - <http://www.gutenberg.spbu.cz/repbase/>

## Genome Sequence Analysis - Step One Need to Mask Repeats

Sequenced  
fragmented  
DNA



Masked DNA  
Sequence



Assembled  
DNA  
Sequence



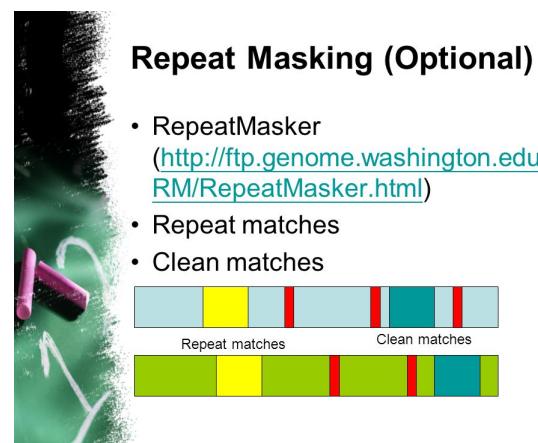
NGS Bioinformatics Workshop -  
1.5 Genome Annotation

## GENOME REPEAT MASKING



MASKED

When a genomes is “masked”, all repeats and low complexity regions are hidden away and replaced with “N”s, so that they will not be aligned to.



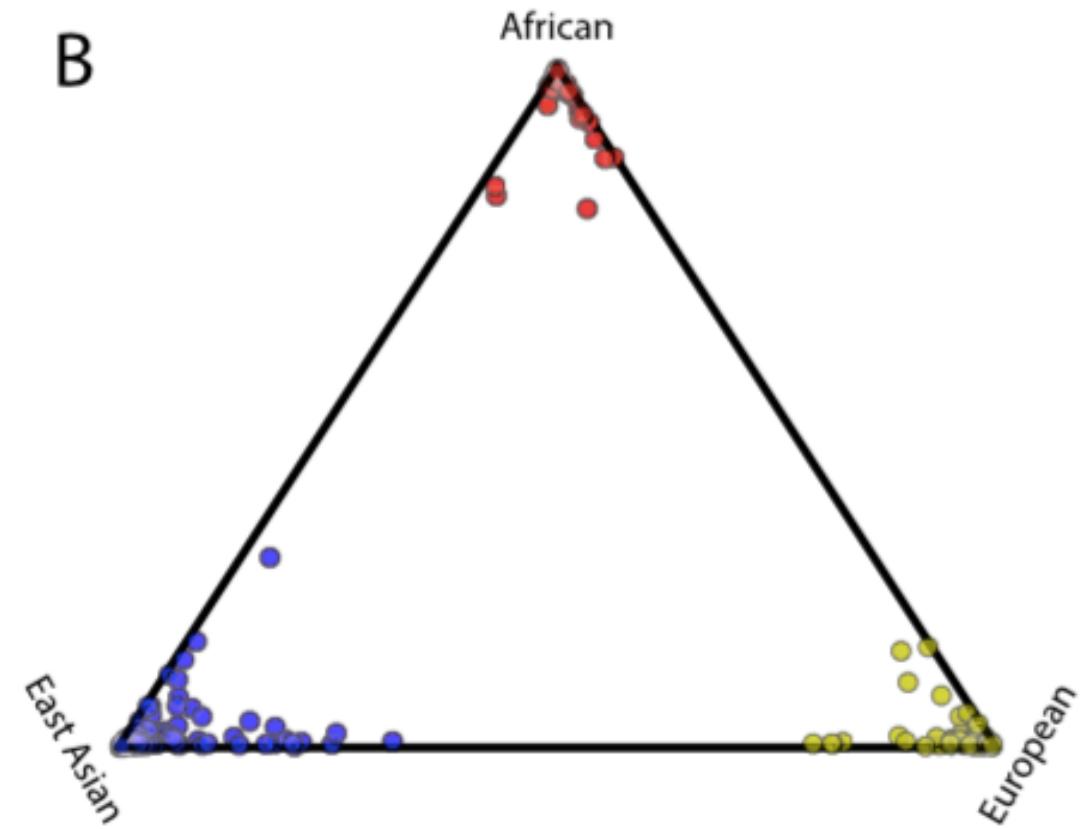
# STR variation in populations

Resource

## The landscape of human STR variation

Thomas Willems,<sup>1,2</sup> Melissa Gymrek,<sup>1,3,4,5</sup> Gareth Highnam,<sup>6</sup>  
The 1000 Genomes Project Consortium, David Mittelman,<sup>6,7</sup> and Yaniv Erlich<sup>1</sup>

~700,000 polymorphic loci  
from 1056 individuals



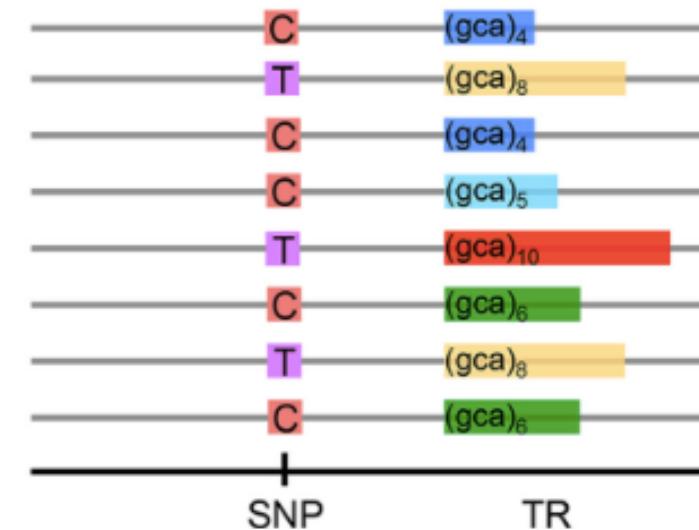
Why should we bother?

-- remember the bliss

# Why should we bother?

-- remember the bliss

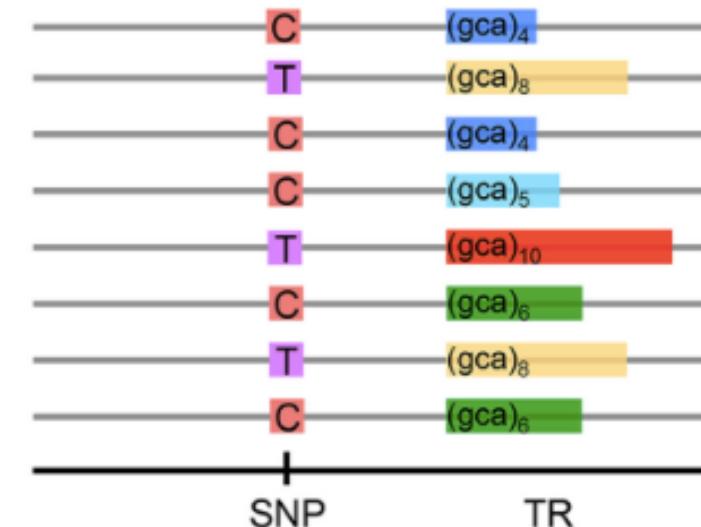
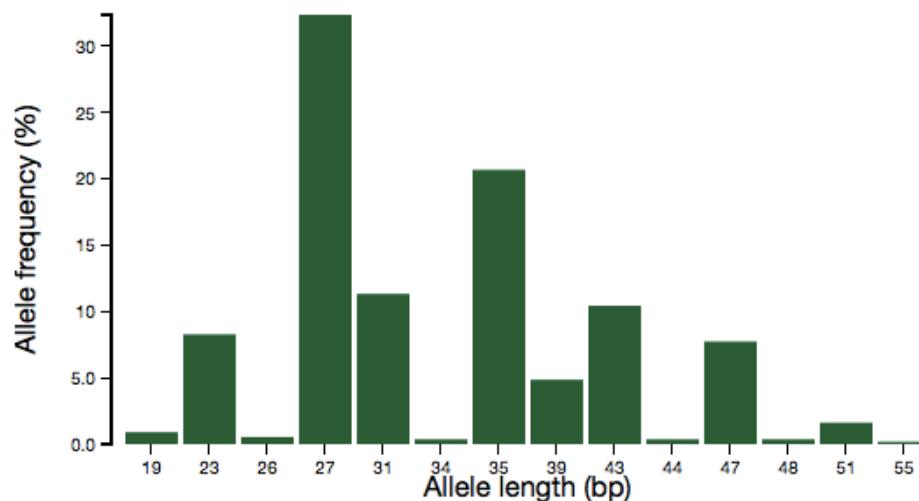
- perhaps greatest number of variants
- much more rapid changes
- not marked by SNPs – no linkage disequilibrium



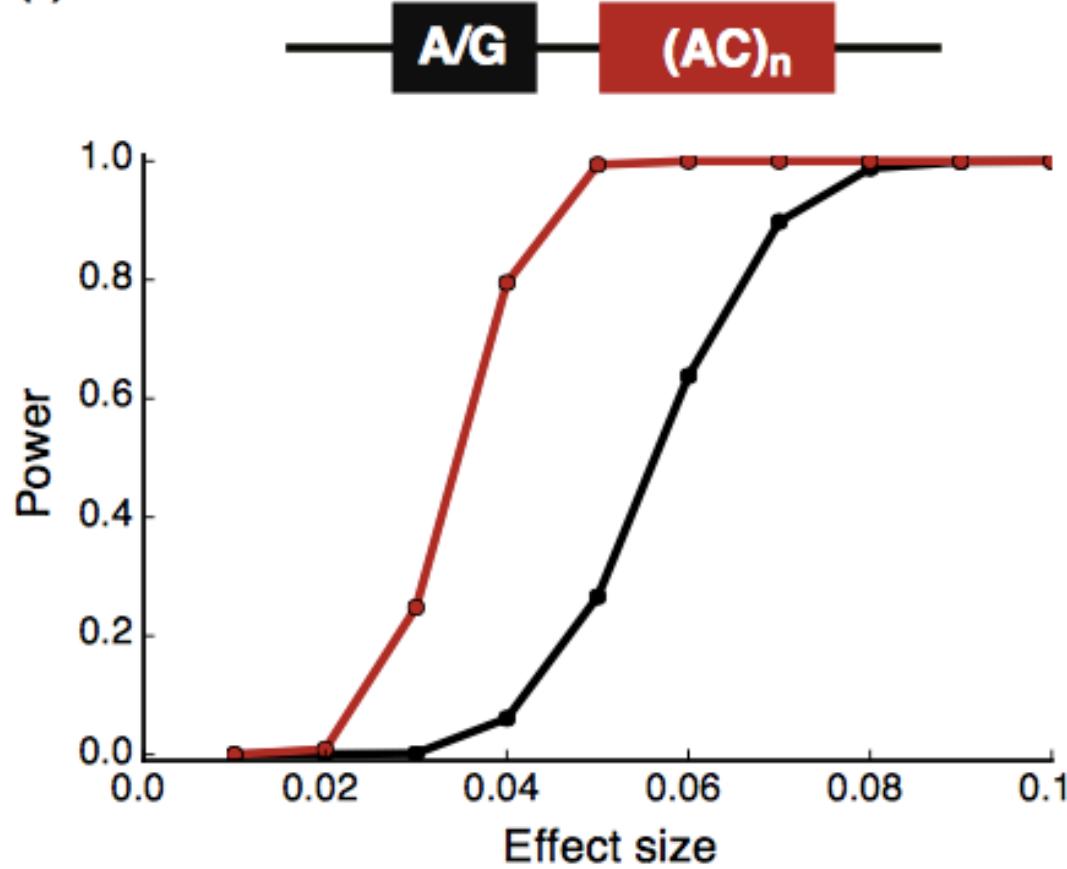
# Why should we bother?

-- remember the bliss

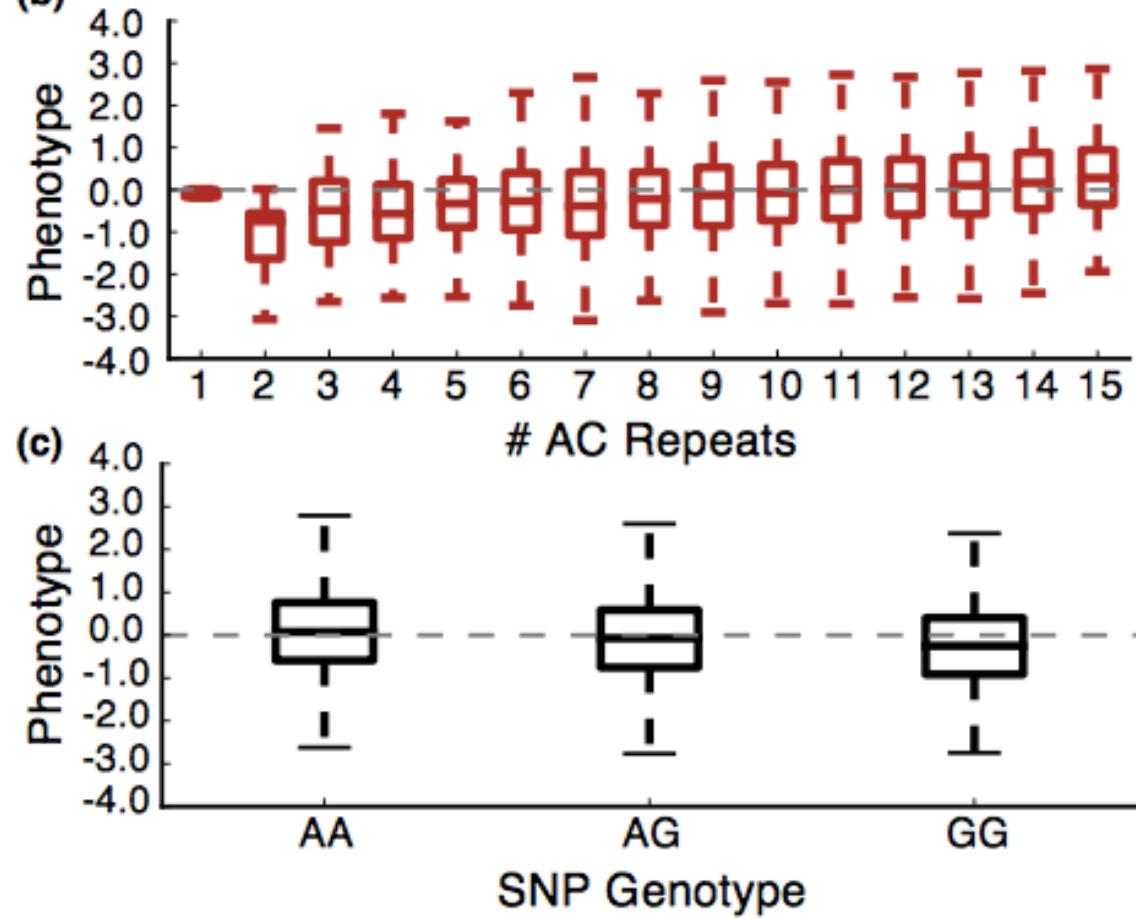
- perhaps greatest number of variants
- much more rapid changes
- not marked by SNPs – no linkage disequilibrium
- continuum of genotypes



(a)



(b)



Current Opinion in Genetics &amp; Development

SNP-based GWAS is underpowered for detecting underlying STR associations.

(a) Power of a causal STR (red) vs. a linked SNP (black) to detect an underlying STR association. An STR with a mutation rate of  $10^{-4}$  mutations per generation was simulated on a SNP haplotype background. A quantitative phenotype was simulated using an additive model in which the phenotype depends on the sum of the two STR alleles at a locus. The plot shows the power of the STR and strongest linked SNP ( $r^2 = 0.37$ ) to detect an association, showing that a single SNP is often underpowered to detect STR associations. (b) and (c) show example relationships between the causal STR (b) and the strongest linked SNP (c) and phenotype. Gray dashed lines denote the average value of the phenotype. See Supplementary Note for simulation details.

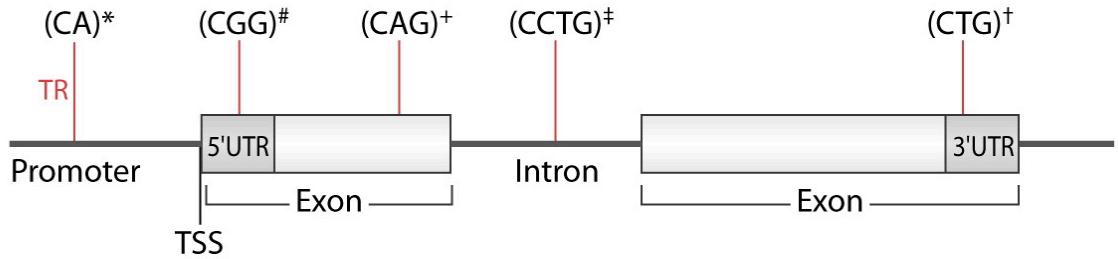
# Why should we bother?

-- remember the bliss

- perhaps greatest number of variants
- much more rapid changes
- not marked by SNPs – no linkage disequilibrium
- continuum of genotypes

Moreover

Impact on gene regulation



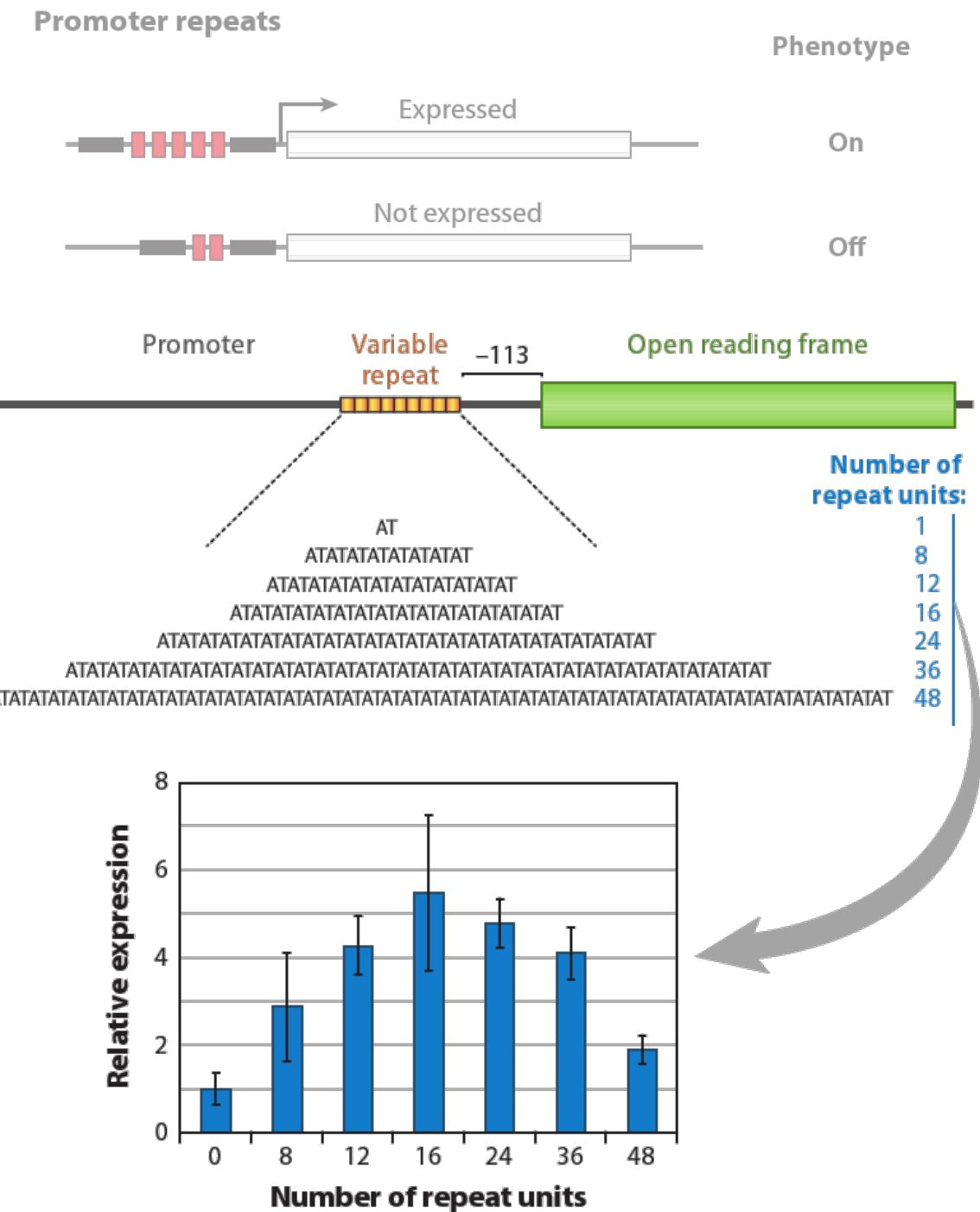
## Microsatellite Tandem Repeats Are Abundant in Human Promoters and Are Associated with Regulatory Elements

Sterling Sawaya<sup>1\*</sup>, Andrew Bagshaw<sup>2</sup>, Emmanuel Buschiazzo<sup>3</sup>, Pankaj Kumar<sup>4</sup>, Shantanu Chowdhury<sup>4,5</sup>, Michael A. Black<sup>6</sup>, Neil Gemmell<sup>1</sup>

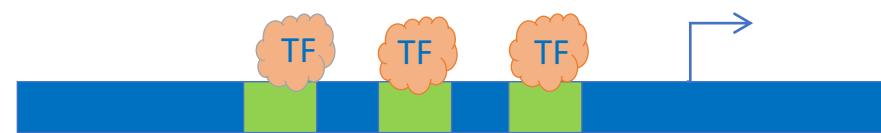
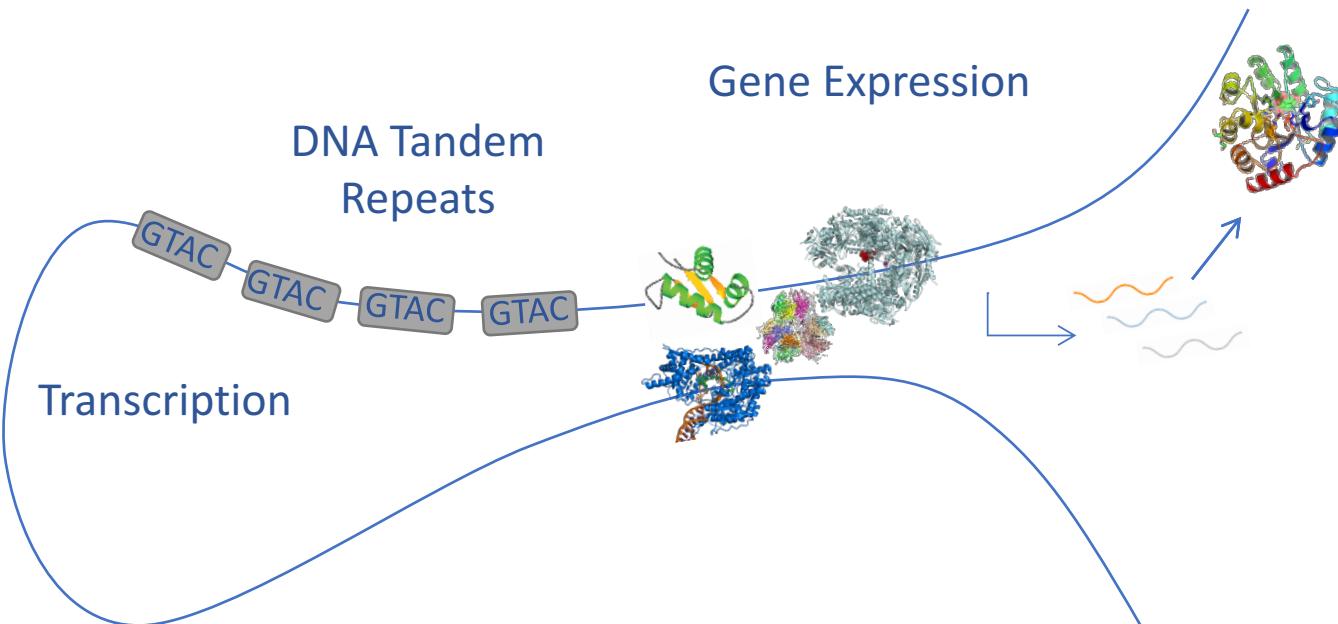
switch mechanism

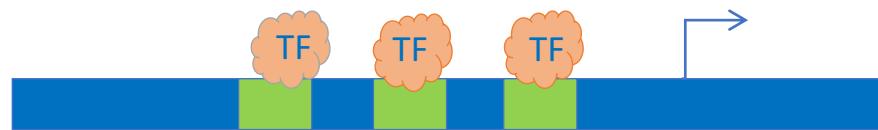
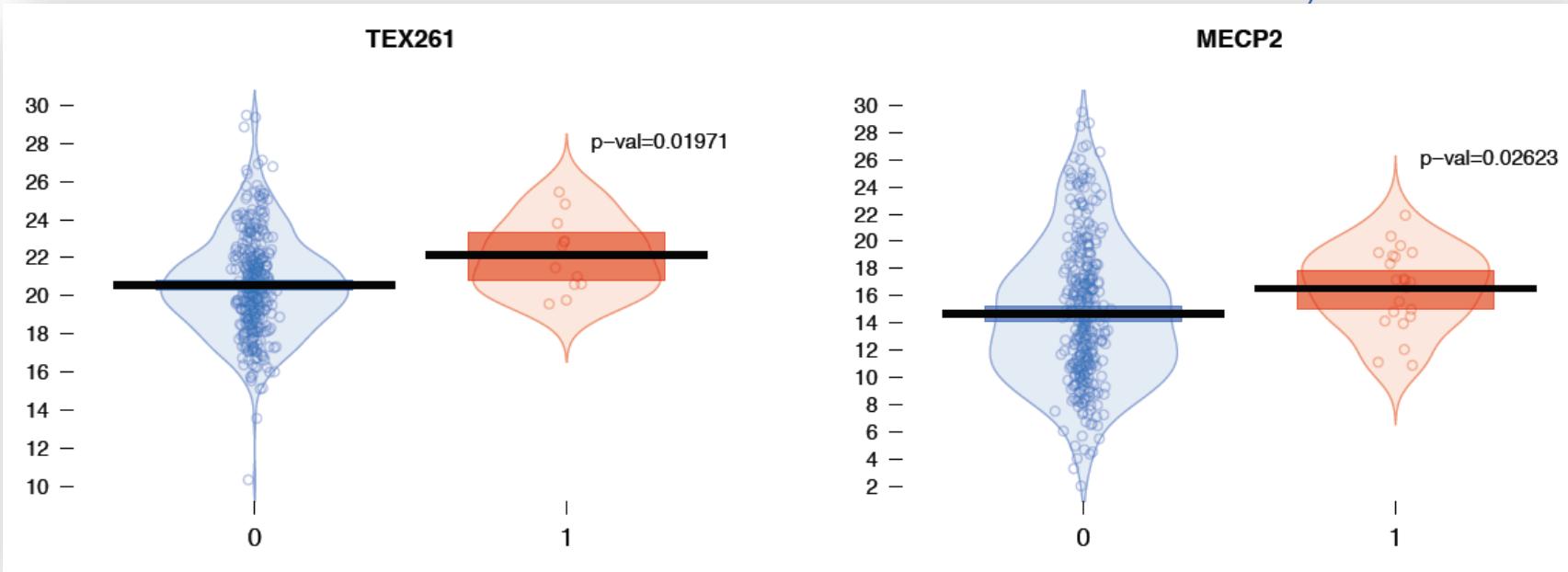


## switch mechanism



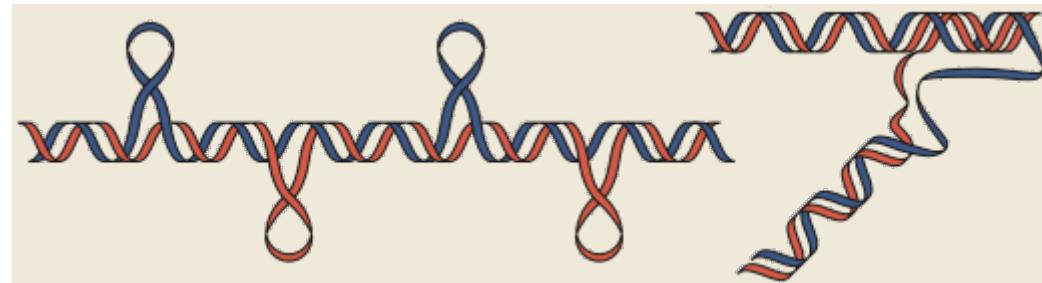
Protein Synthesis



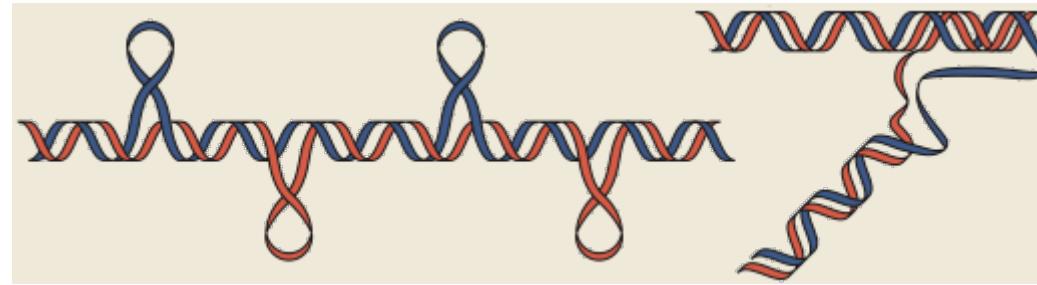


Bilgin et al, in prep

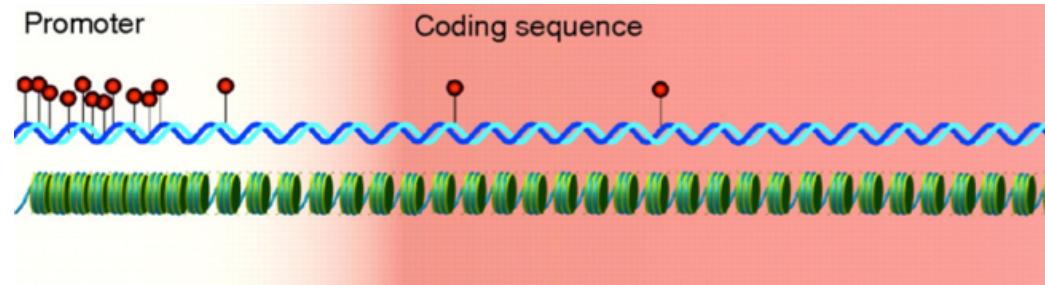
form secondary DNA  
structures



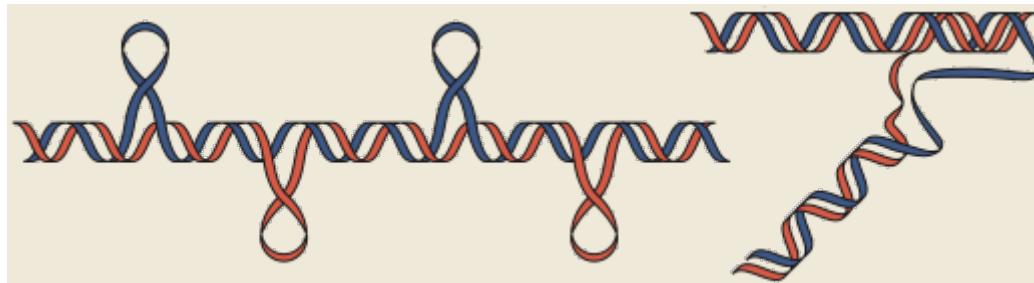
form secondary DNA  
structures



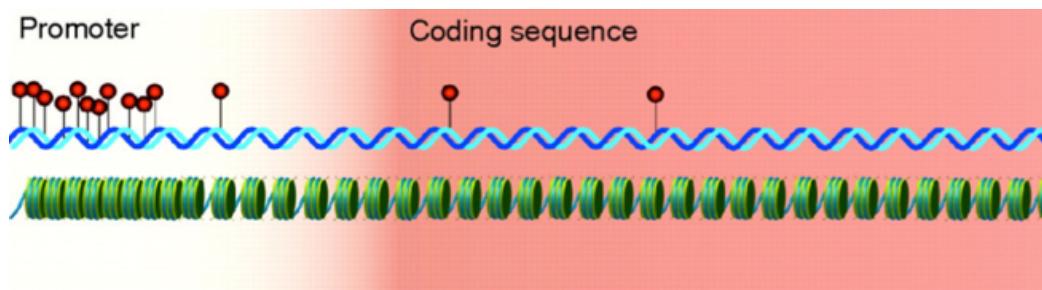
mark methylation  
sites



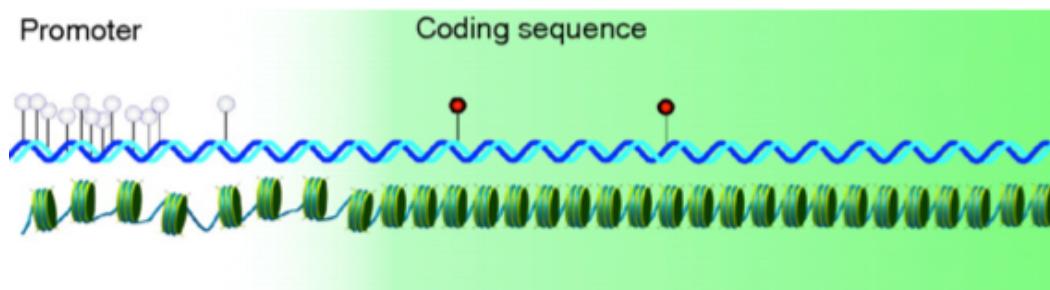
form secondary DNA structures



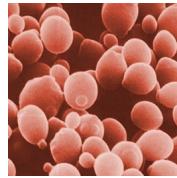
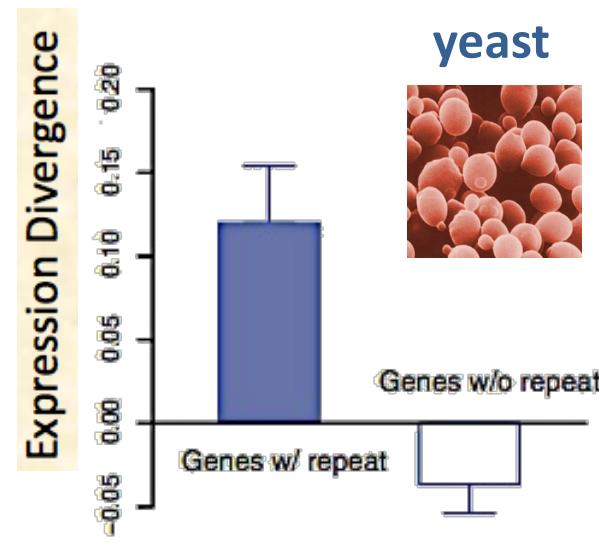
mark methylation sites



avoid nucleosome formation



yeast



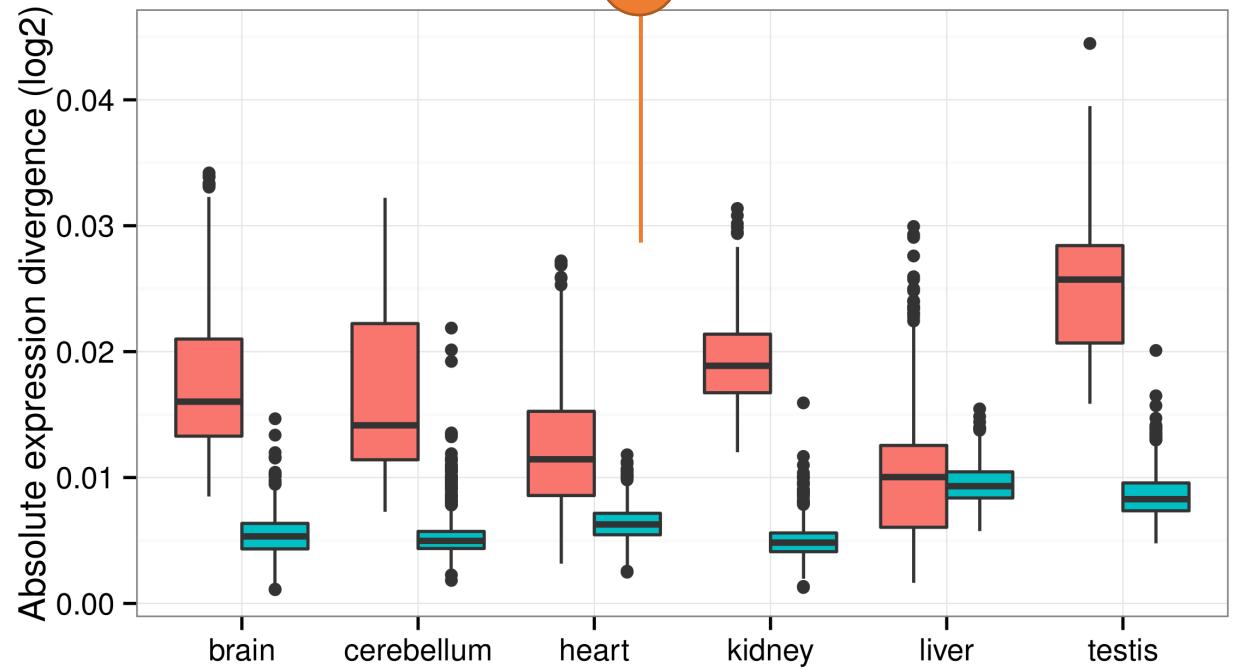
Vinces *et al.*, *Science*, 2009



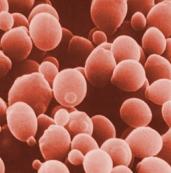
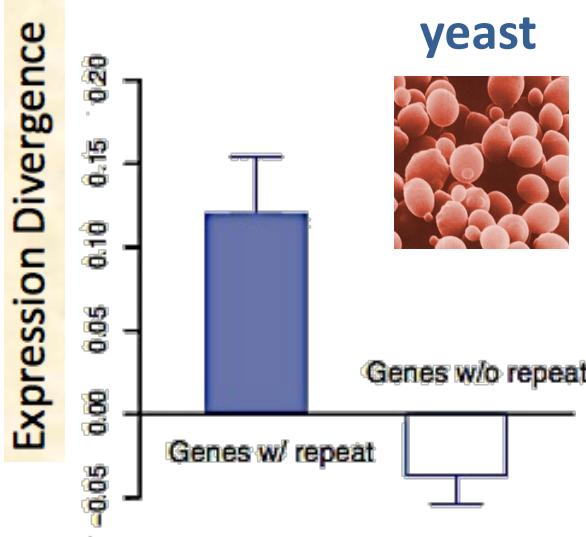
macaque

chimp

human



TR  
with  
without



yeast

Vinces *et al.*, *Science*, 2009

Bilgin Sonay *et al.*, *Genome Res*, 2015

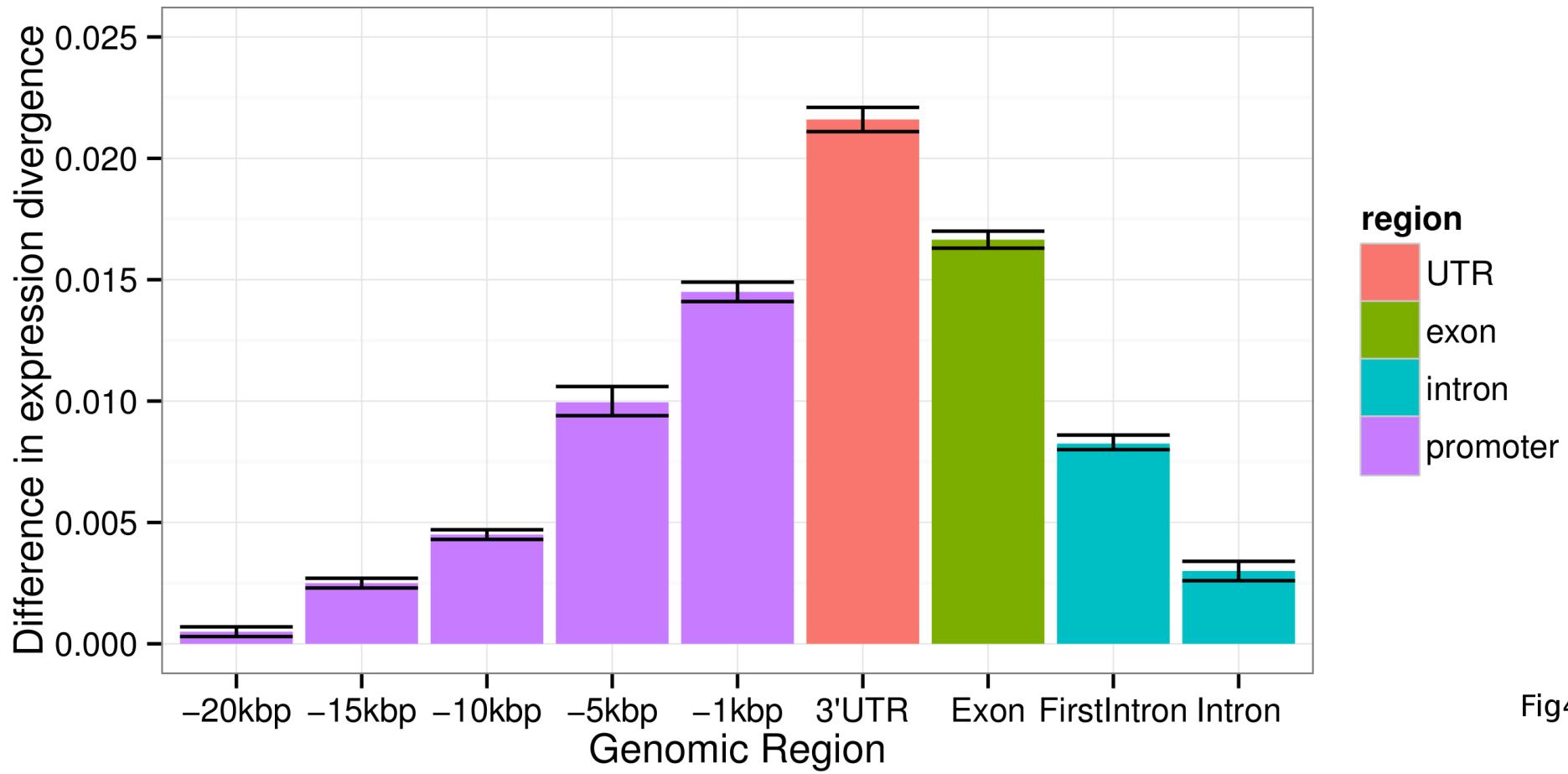


Fig4

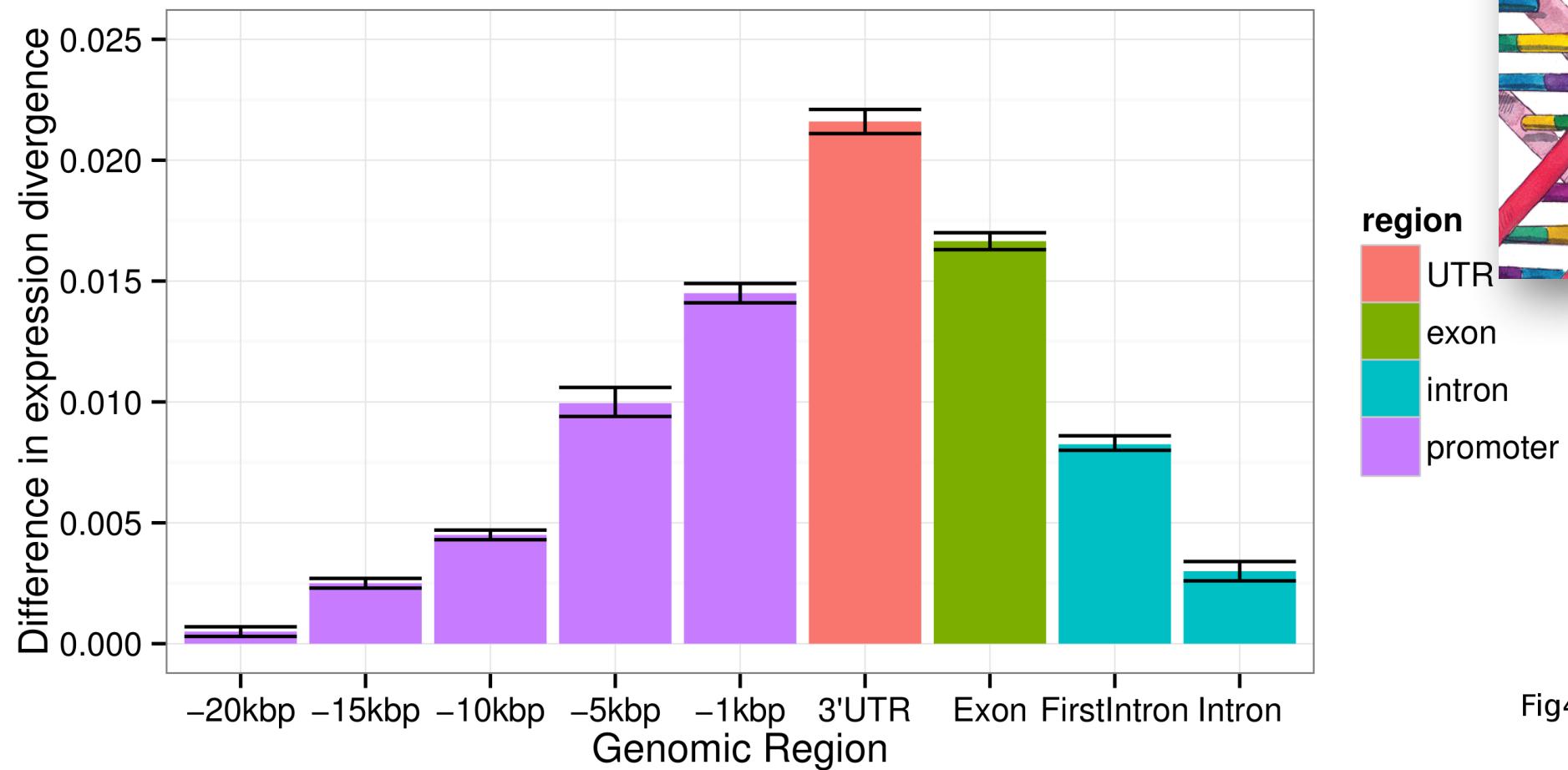
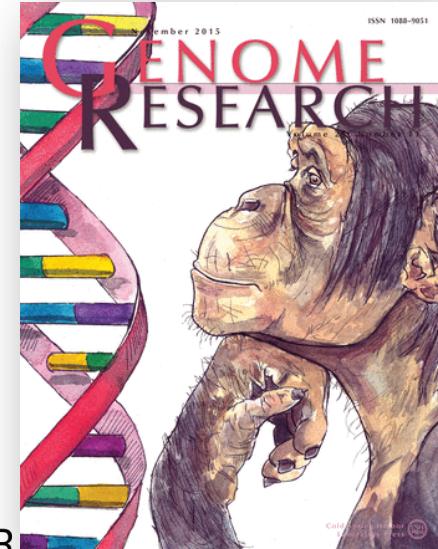


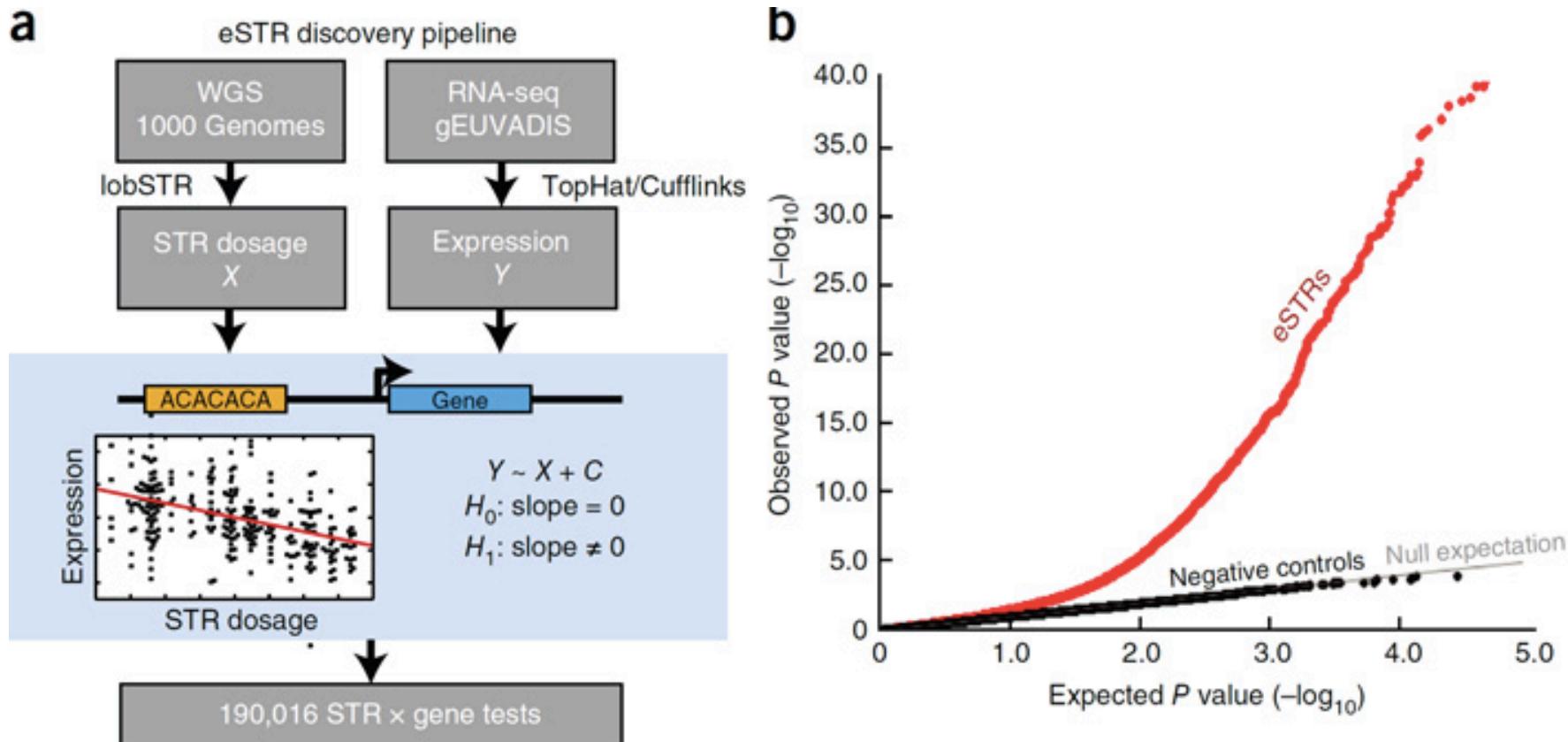
Fig4

## Abundant contribution of short tandem repeats to gene expression variation in humans

Melissa Gymrek<sup>1-4</sup>, Thomas Willems<sup>1,4,5</sup>, Audrey Guilmatre<sup>6,7</sup>, Haoyang Zeng<sup>8</sup>, Barak Markus<sup>1</sup>, Stoyan Georgiev<sup>9</sup>, Mark J Daly<sup>3,10</sup>, Alkes L Price<sup>3,11,12</sup>, Jonathan K Pritchard<sup>9,13</sup>, Andrew J Sharp<sup>6</sup> & Yaniv Erlich<sup>1,4,14,15</sup>

- 10-15% impact on gene expression

- 2060 genes effected by eSTRs



# Why should we bother?

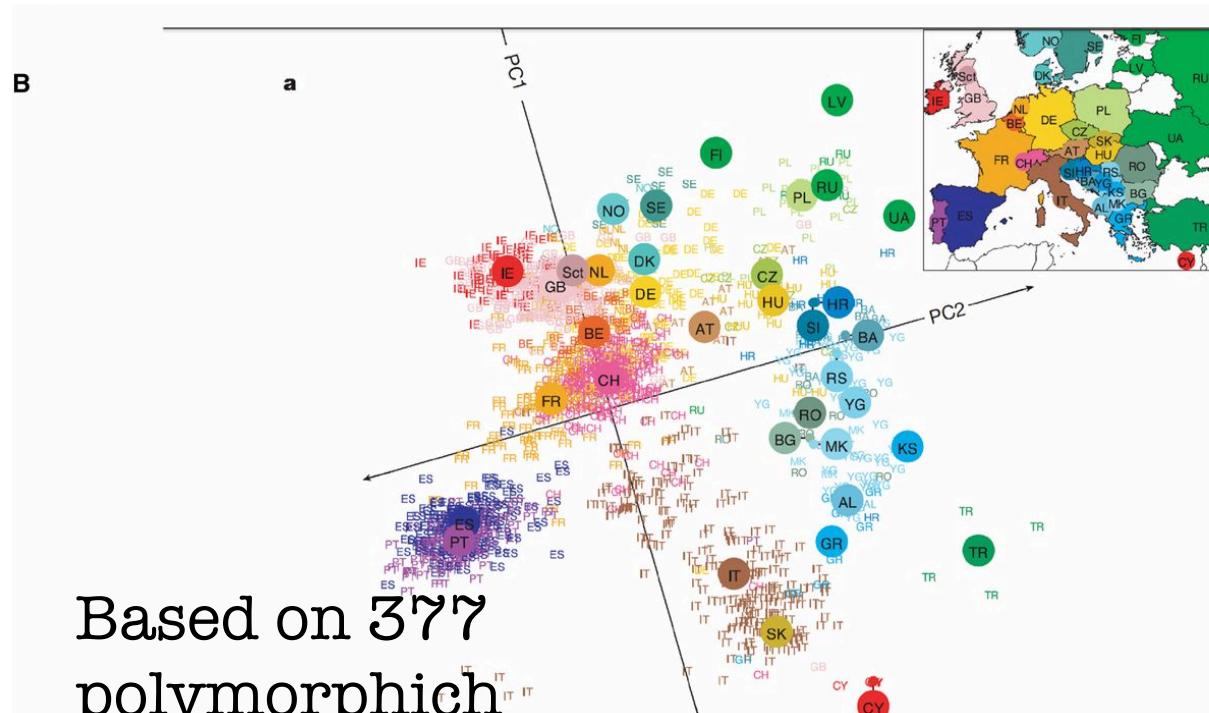
-- remember the bliss

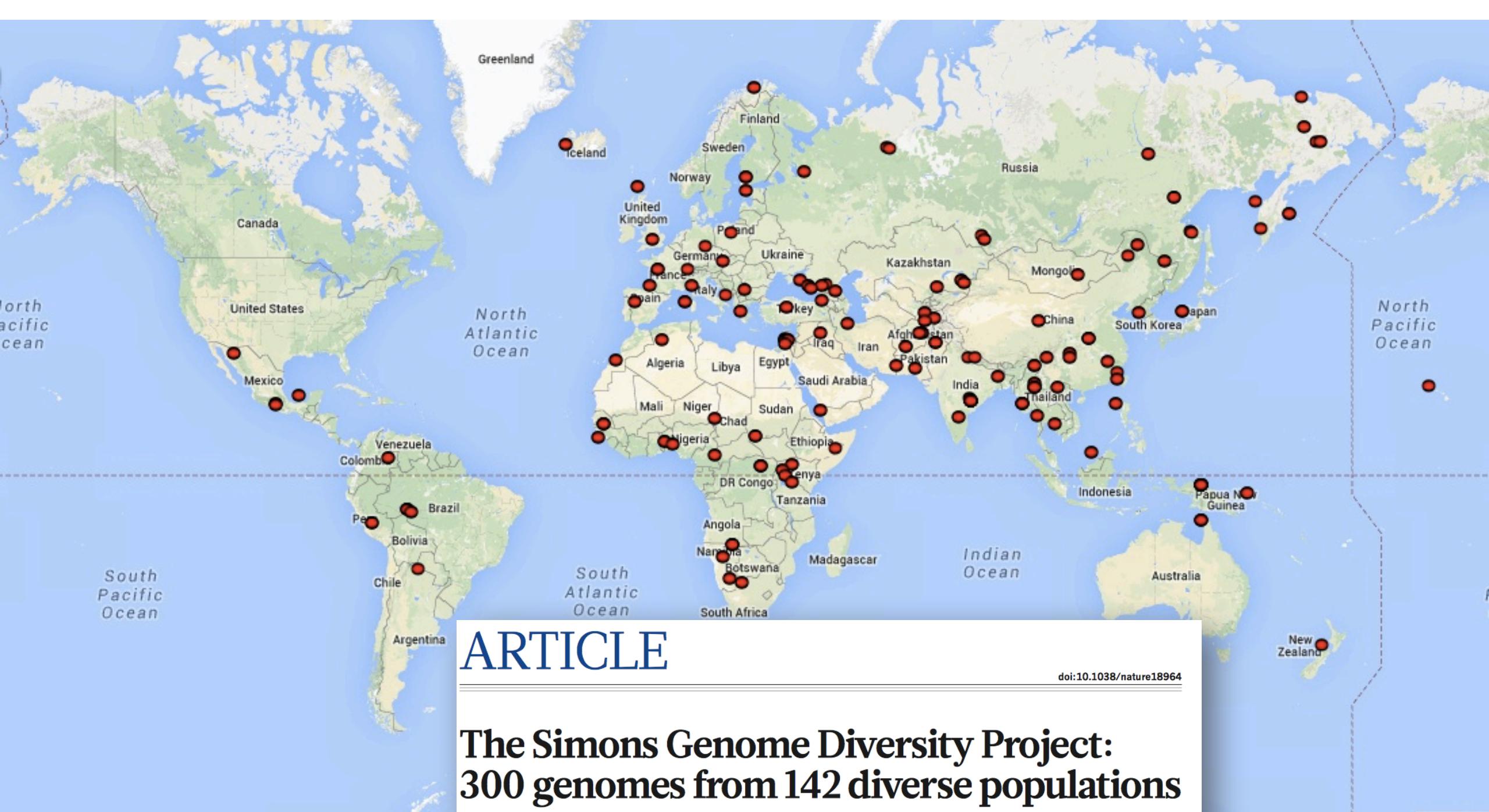
- perhaps greatest number of variants
- much more rapid changes
- not marked by SNPs – no linkage disequilibrium
- continuum of genotypes

**Moreover**

Impact on gene regulation

Greater resolution at recent or within species changes



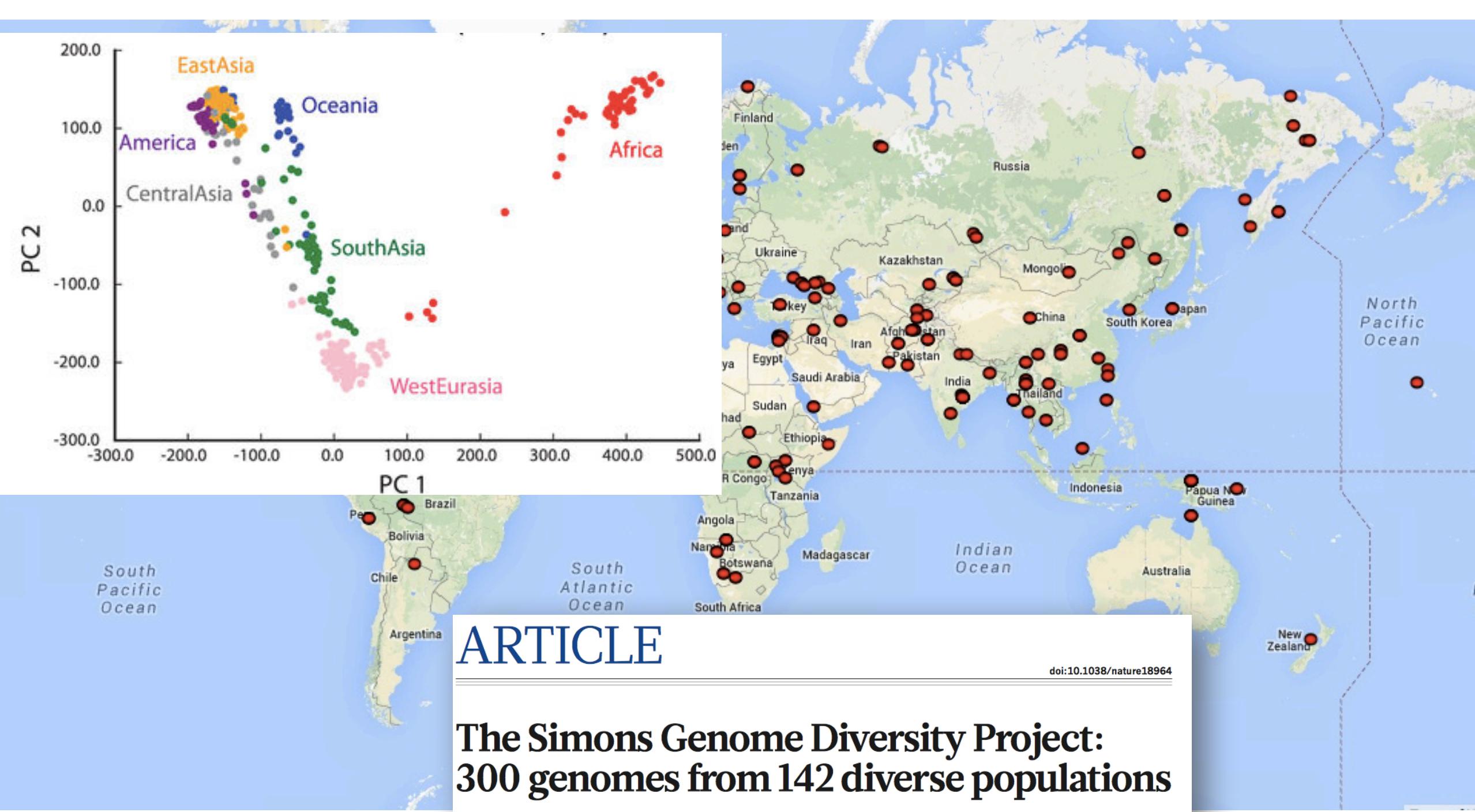


## ARTICLE

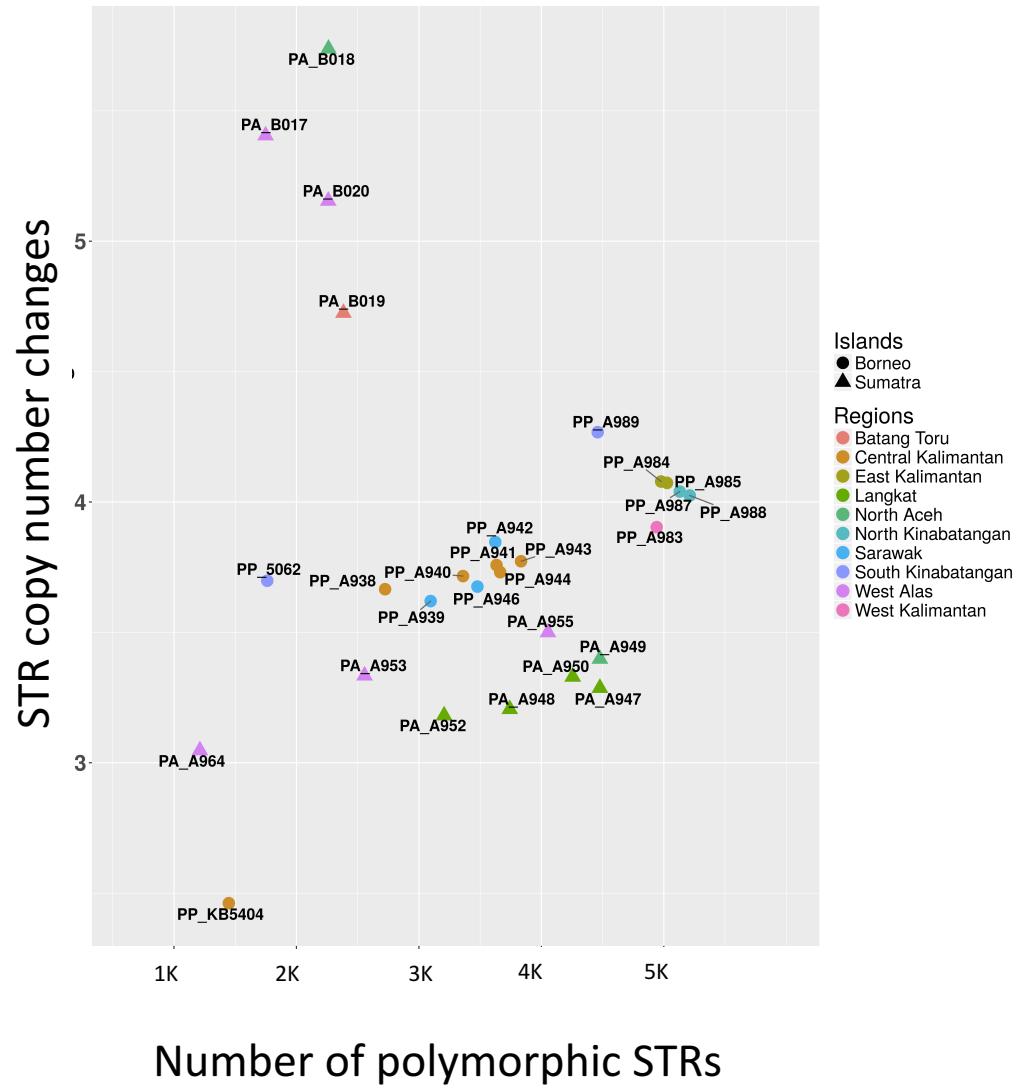
doi:10.1038/nature18964

# The Simons Genome Diversity Project: 300 genomes from 142 diverse populations

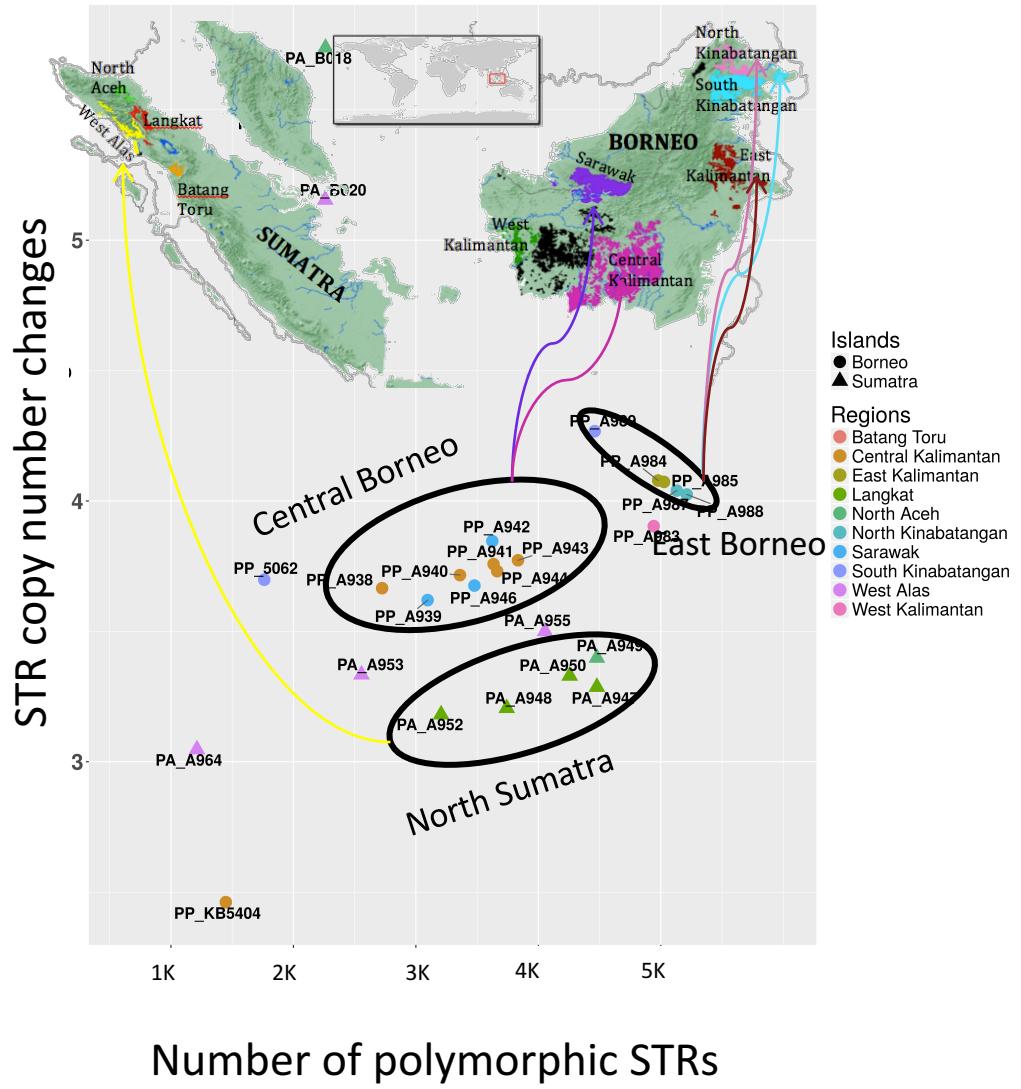
A list of authors and affiliations appears at the end of the paper.



# Evolution of Recent populations

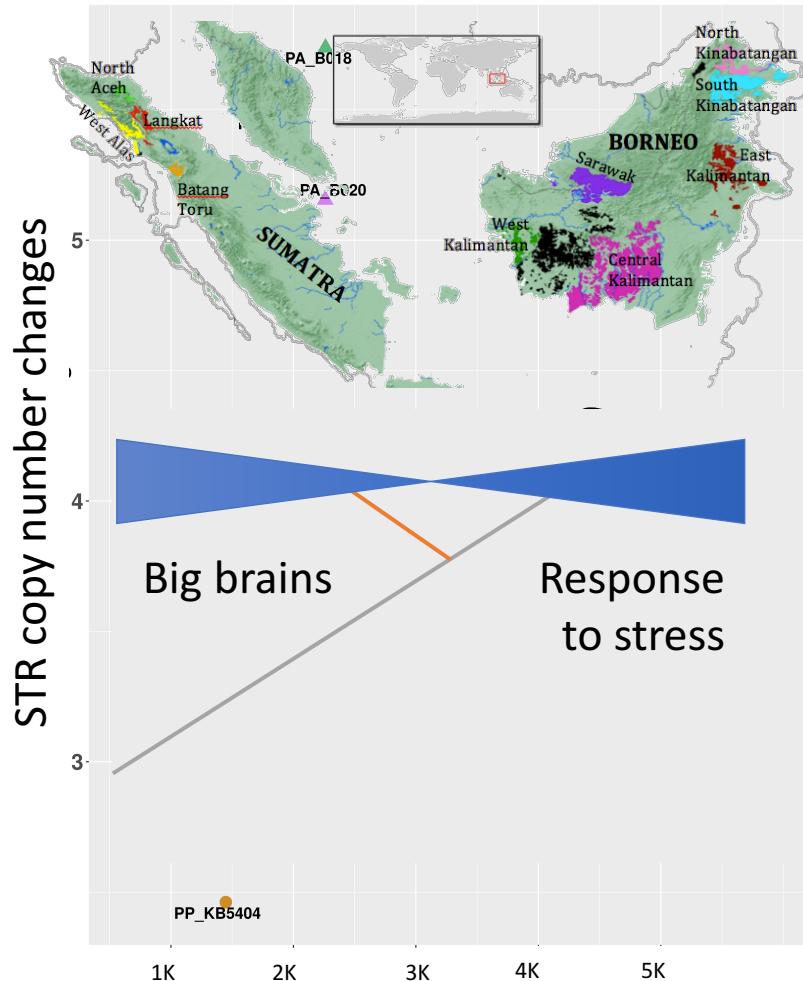


# Evolution of Recent populations



Voici, Kruetzen & Bilgin Sonay in prep.

# Evolution of Recent populations



Number of polymorphic STRs



- Islands  
● Borneo  
▲ Sumatra
- Regions  
● Batang Toru  
● Central Kalimantan  
● East Kalimantan  
● Langkat  
● North Aceh  
● North Kinabatangan  
● Sarawak  
● South Kinabatangan  
● West Alas  
● West Kalimantan



# Why should we bother?

# -- remember the bliss

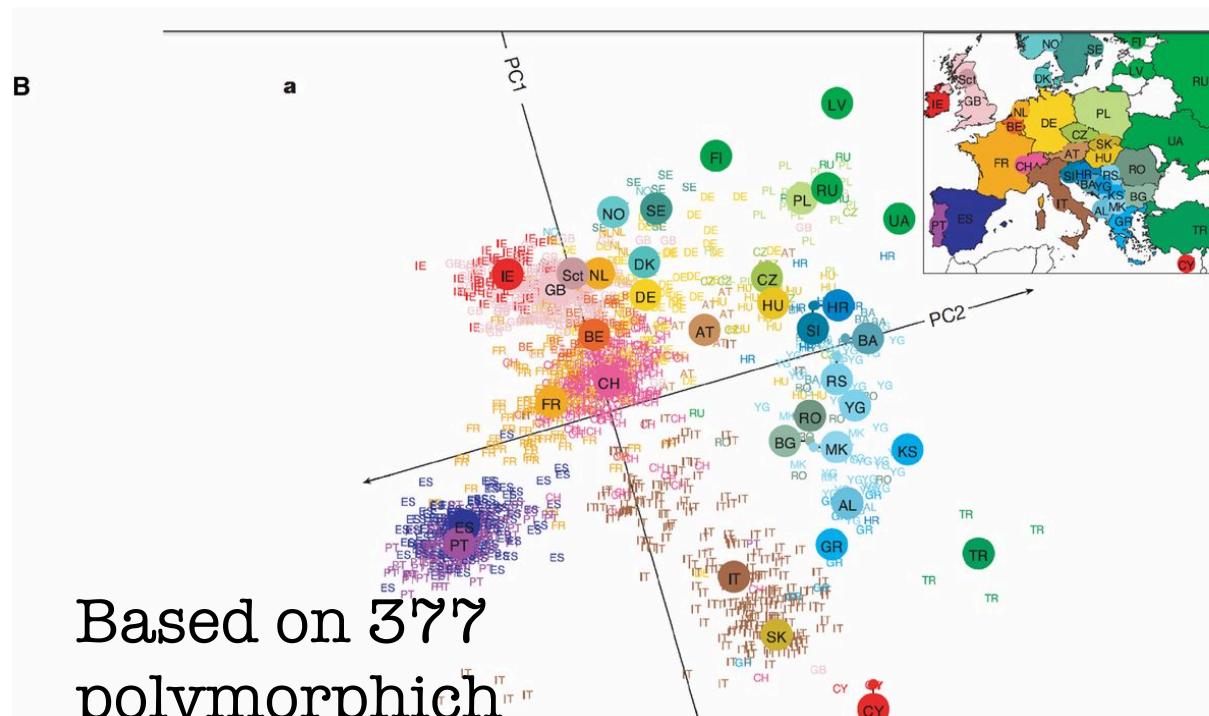
- perhaps greatest number of variants
  - much more rapid changes
  - not marked by SNPs – no linkage disequilibrium
  - continuum of genotypes

## Moreover

## **Impact on gene regulation**

**Greater resolution at recent or within species changes**

# Selection, adaptation



Based on 377  
polymorphich  
STR loci

Rosenberg *et al.* 2002

## maternity in voles

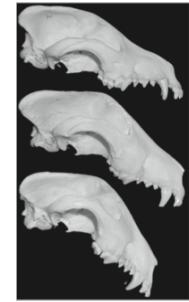


Gamayel *et al.*,  
*Annu. Rev. Genet.*,  
2010

**maternity in voles**



**dog skulls**

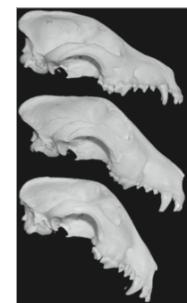


Gamayel *et al.*,  
*Annu. Rev. Genet.*,  
2010

**maternity in voles**



**dog skulls**



**adaptation**

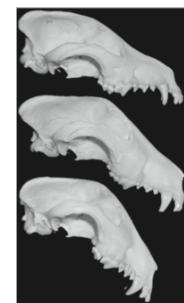


Gamayel *et al.*,  
*Annu. Rev. Genet.*,  
2010

**maternity in voles**



**dog skulls**



**adaptation**



**survival**



Gamayel *et al.*,  
*Annu. Rev. Genet.*,  
2010

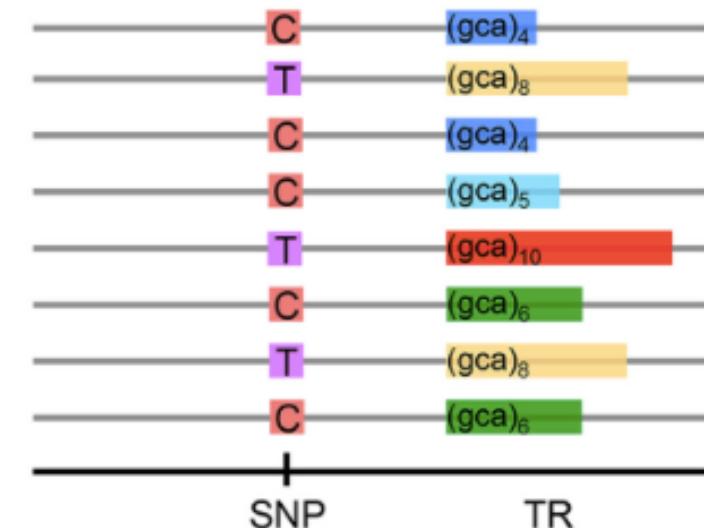
# Why should we bother?

-- remember the bliss

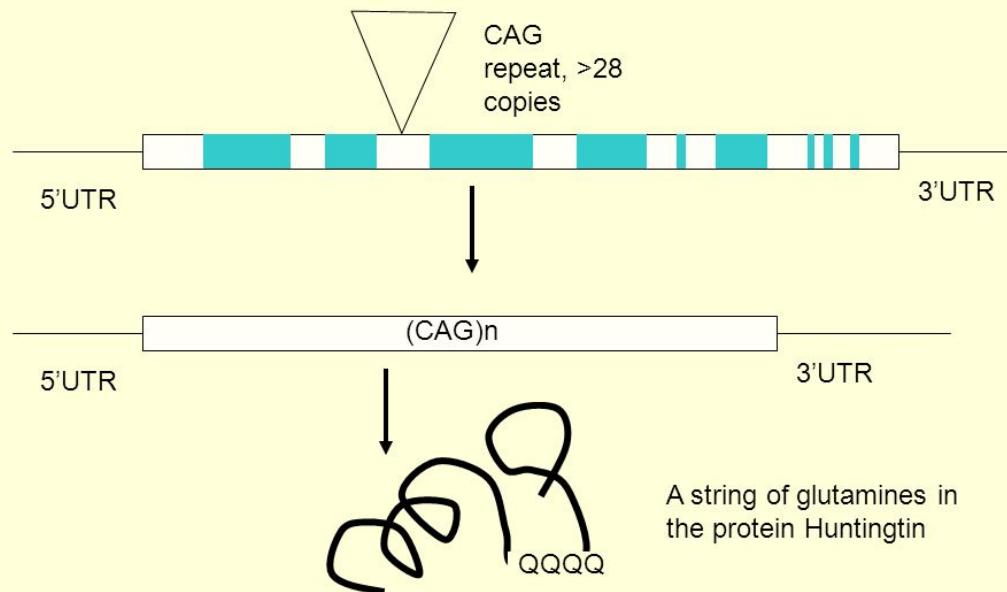
- perhaps greatest number of variants
- much more rapid changes
- not marked by SNPs – no linkage disequilibrium

**Moreover**

is involved in gene regulation, recent evolution, adaptation  
is involved in disease

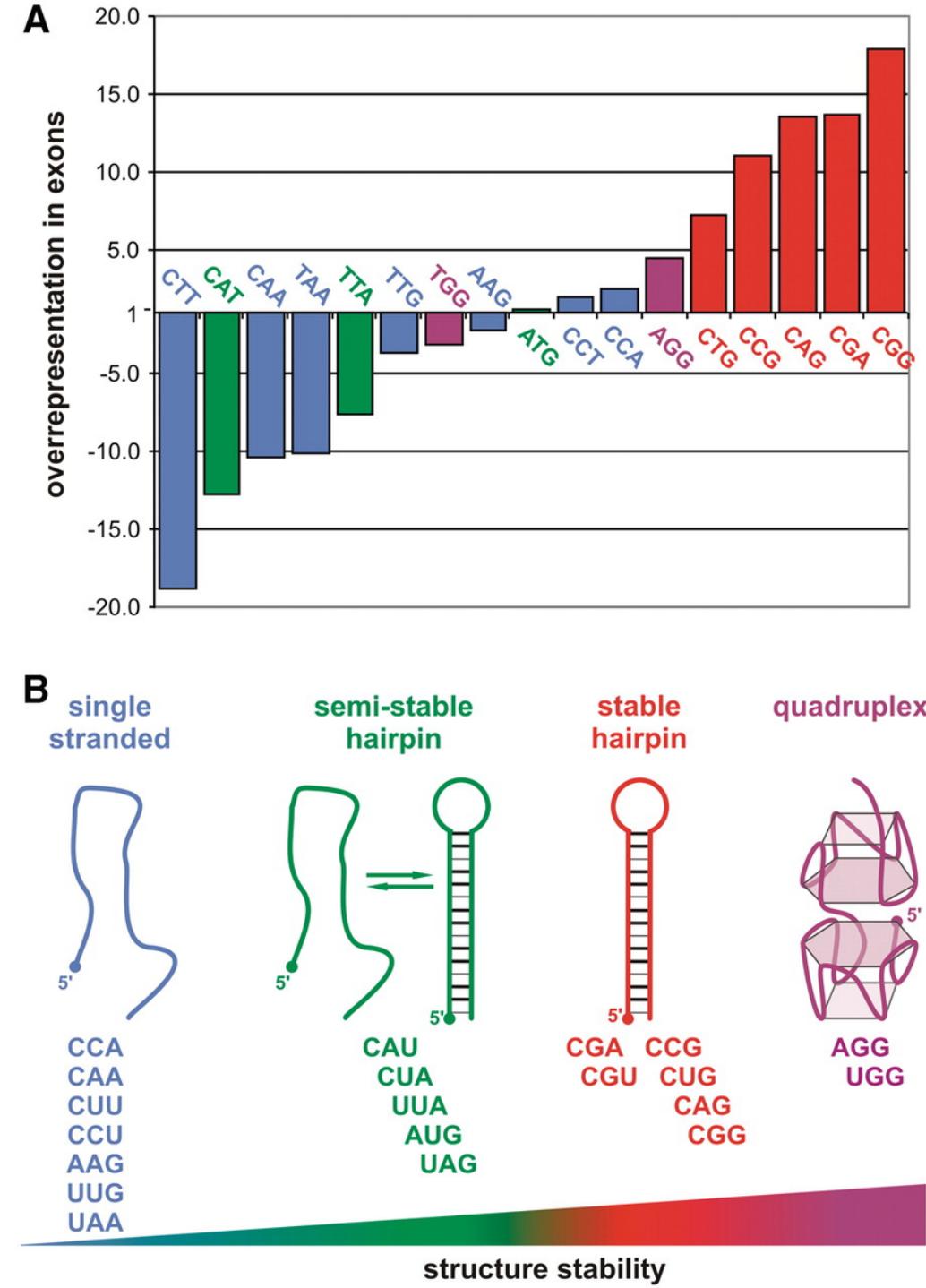


# Huntington's Disease



# Huntington's Disease

The diagram illustrates the expansion of a CAG triplet repeat in the 3' UTR of the huntingtin gene. In the top panel, a normal gene structure is shown with a 5' UTR (white) and a 3' UTR (teal). A triangle above the 3' UTR indicates a CAG repeat of 28 or fewer copies. In the bottom panel, an arrow points down to a mutated gene where the CAG repeat has expanded to (CAG)<sub>n</sub>, where n is greater than 28. This expanded region is shown as a long white box. An arrow from this box points down to a tangled protein structure labeled "QQQQ". To the right, the text reads: "A string of glutamines in the protein Huntingtin".

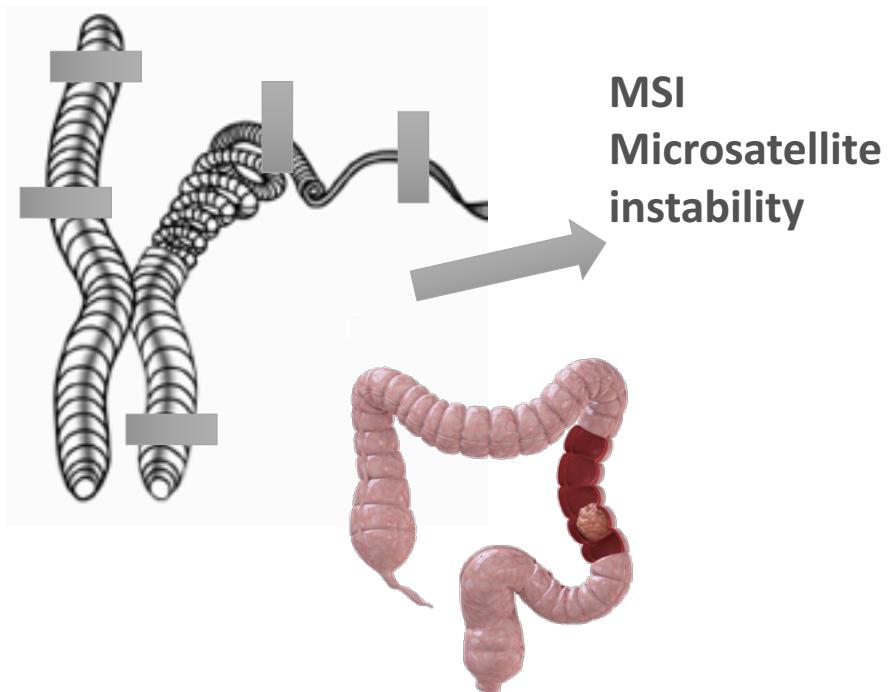


# Colorectal Cancer

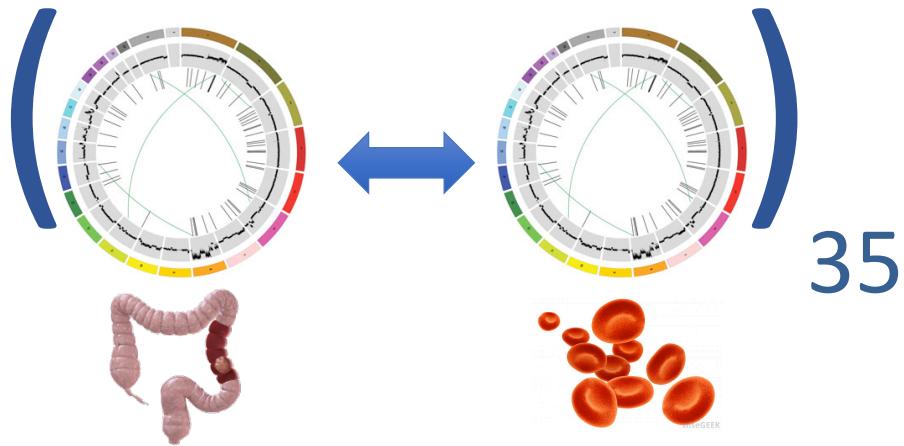
**3<sup>rd</sup> most common in the world**

**2<sup>nd</sup> most lethal in western countries**

Bethesda Guidelines  
for diagnosis

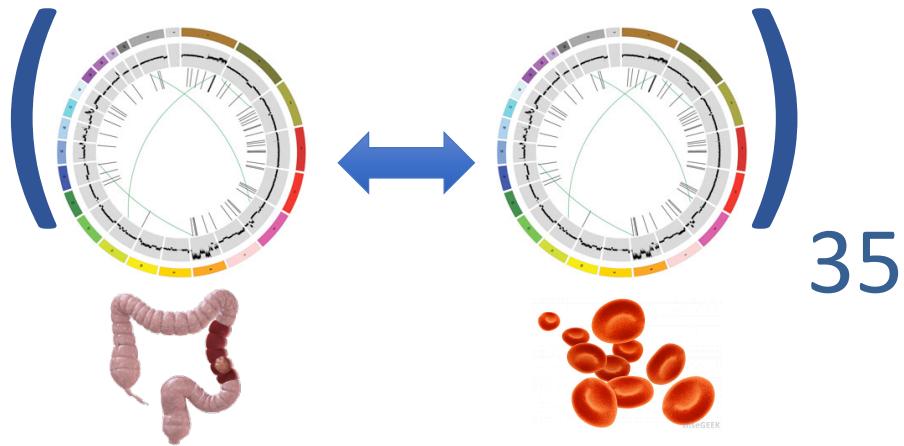


- Promoters (5 kb upstream): Whole genome Sequences
- Exons: Exome Sequences



- RNA-seq for >18,000 genes in 35 tumors

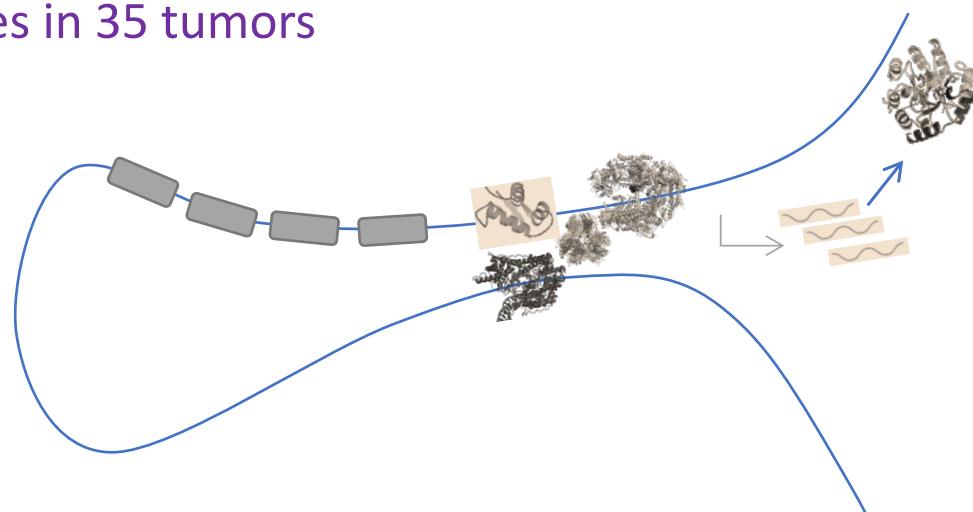
- Promoters (5 kb upstream): Whole genome Sequences
  - Exons: Exome Sequences



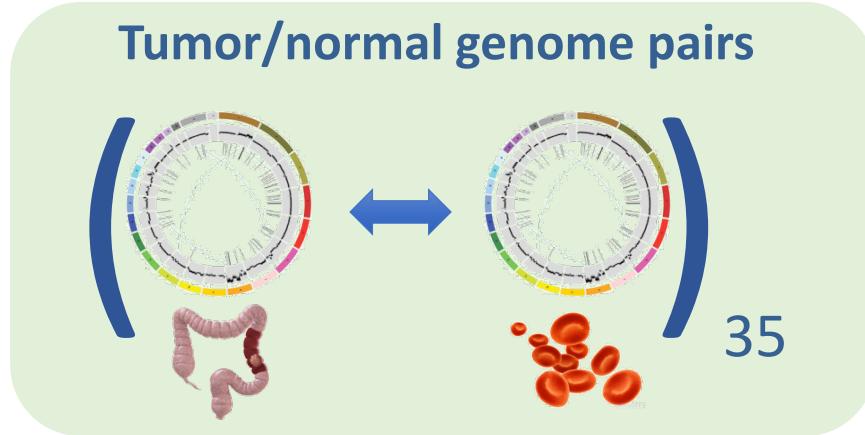
- RNA-seq for >18,000 genes in 35 tumors

# TANDEM REPEATS FINDER

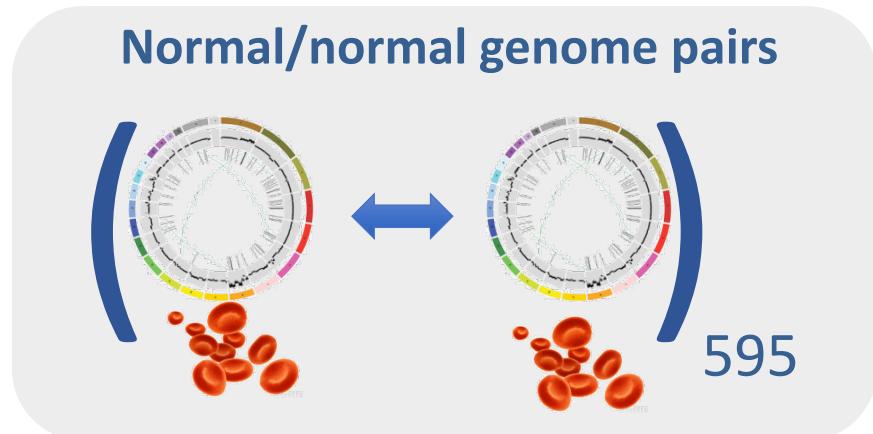
- high purity repeats  
units < 100 nucleotides



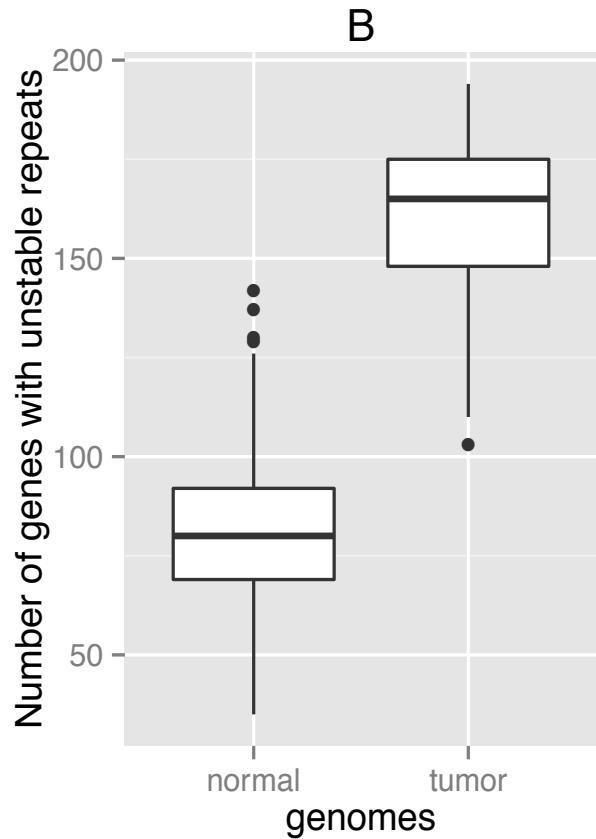
# Colorectal Cancer



Control Set:



A two-fold increase in  
repeat instability in  
cancer promoters

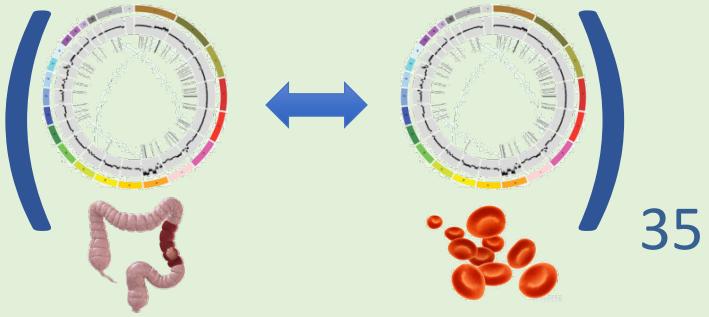


Bilgin Sonay et al., BMC Genomics, 2015

# Colorectal Cancer

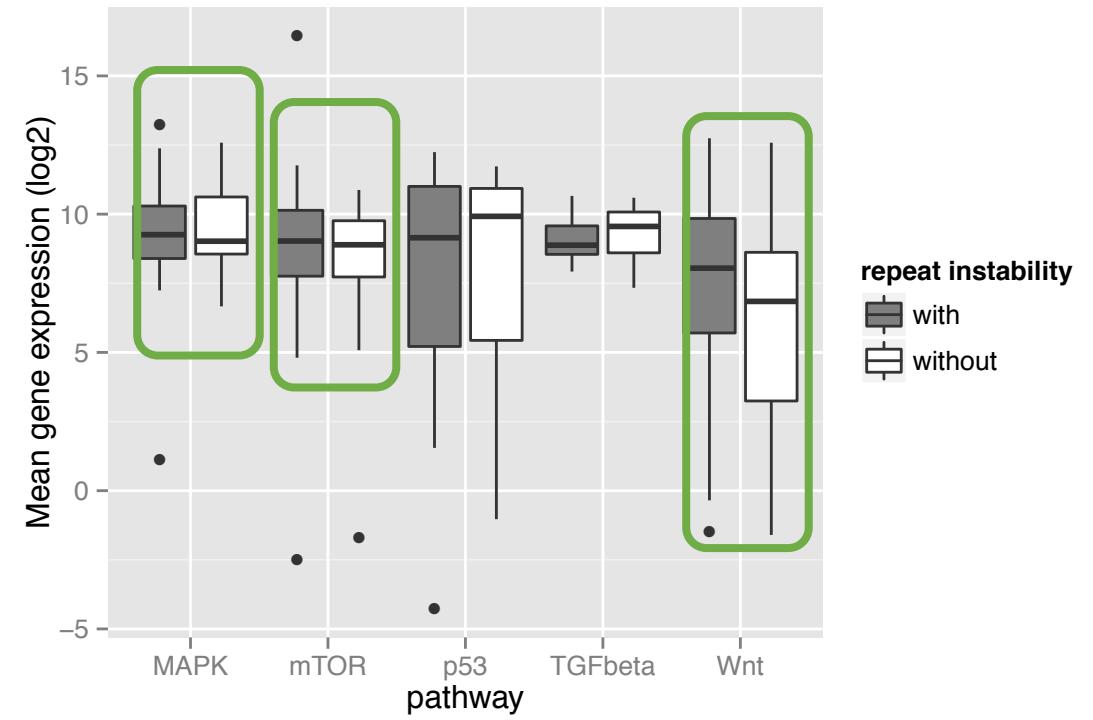
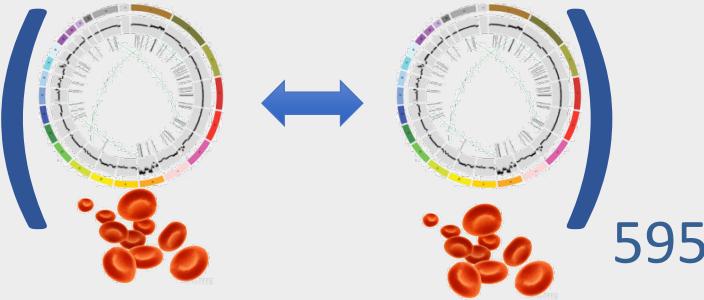
Genes in oncogenic pathways  
with repeat instability have  
increased expression

Tumor/normal genome pairs

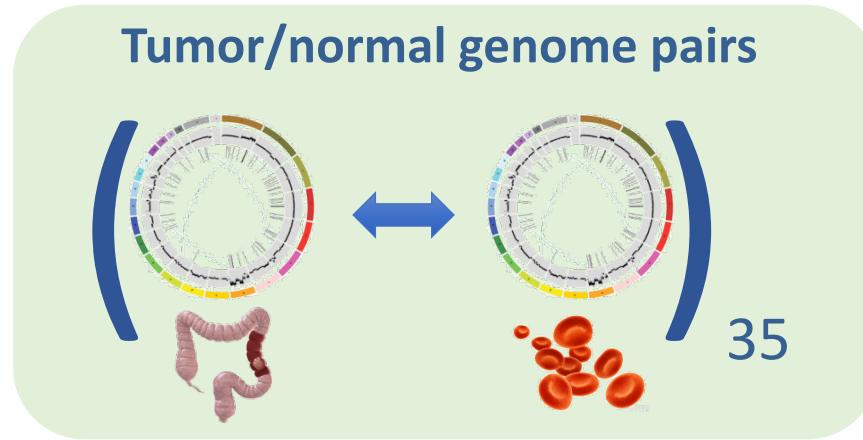


Control Set:

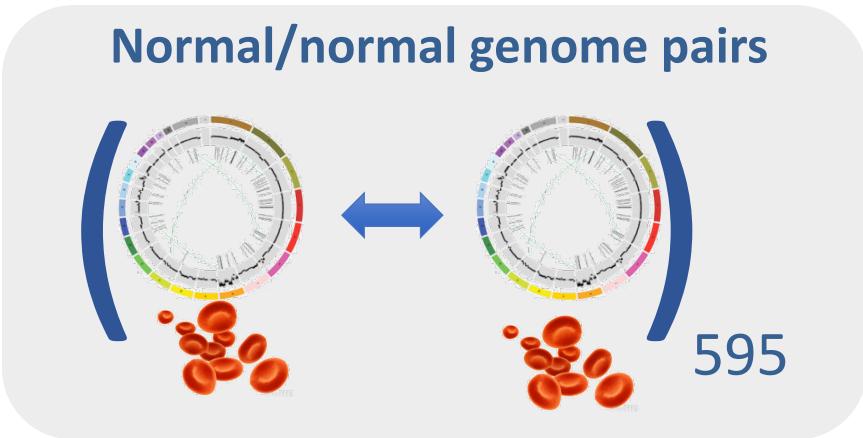
Normal/normal genome pairs



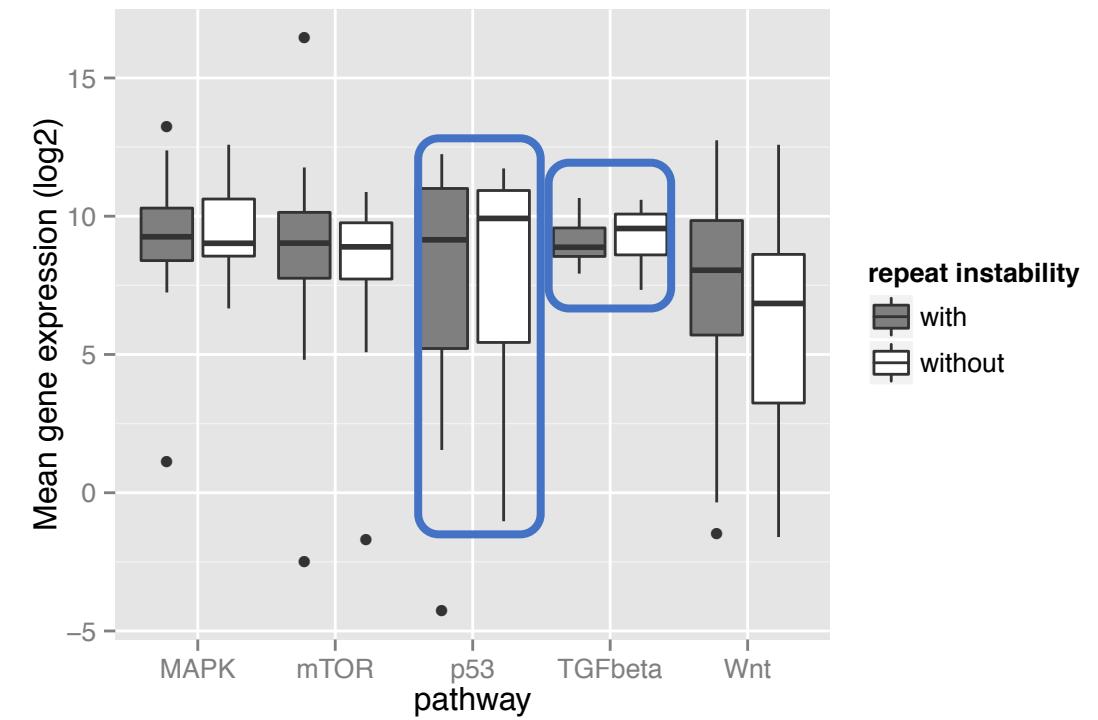
# Colorectal Cancer



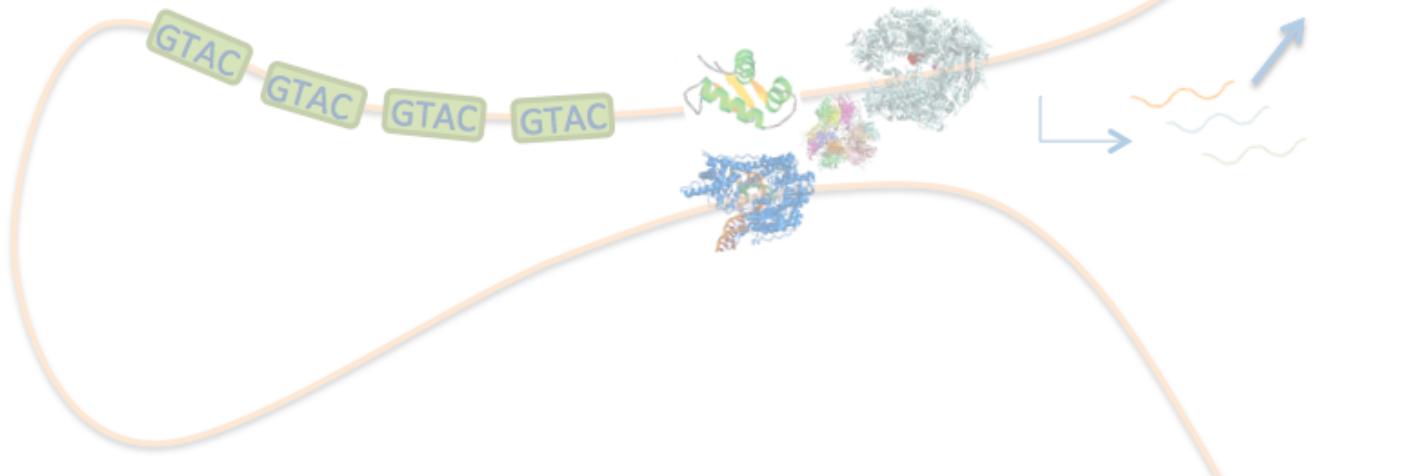
Control Set:



Genes in tumor suppressor pathways with repeat instability have decreased expression



tandem repeat instability ↔ differential gene regulation ↔ disease



tandem repeat  
instability

differential  
gene regulation

disease

collaborators:

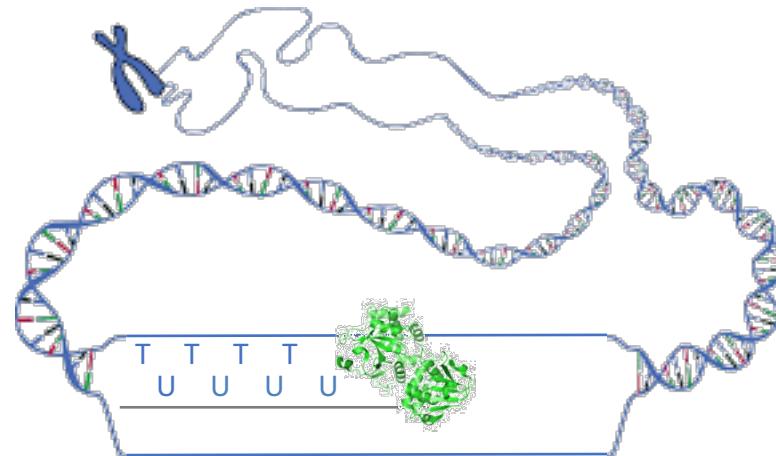
Andreas Wagner  
Tiago Carvalho  
Tomas Bonnet  
Mark Robinson

Andreas Wagner  
Elina Koletou

Joshua Payne  
Stephany Orjuela

Michael Kruetzen  
Alina Voicu

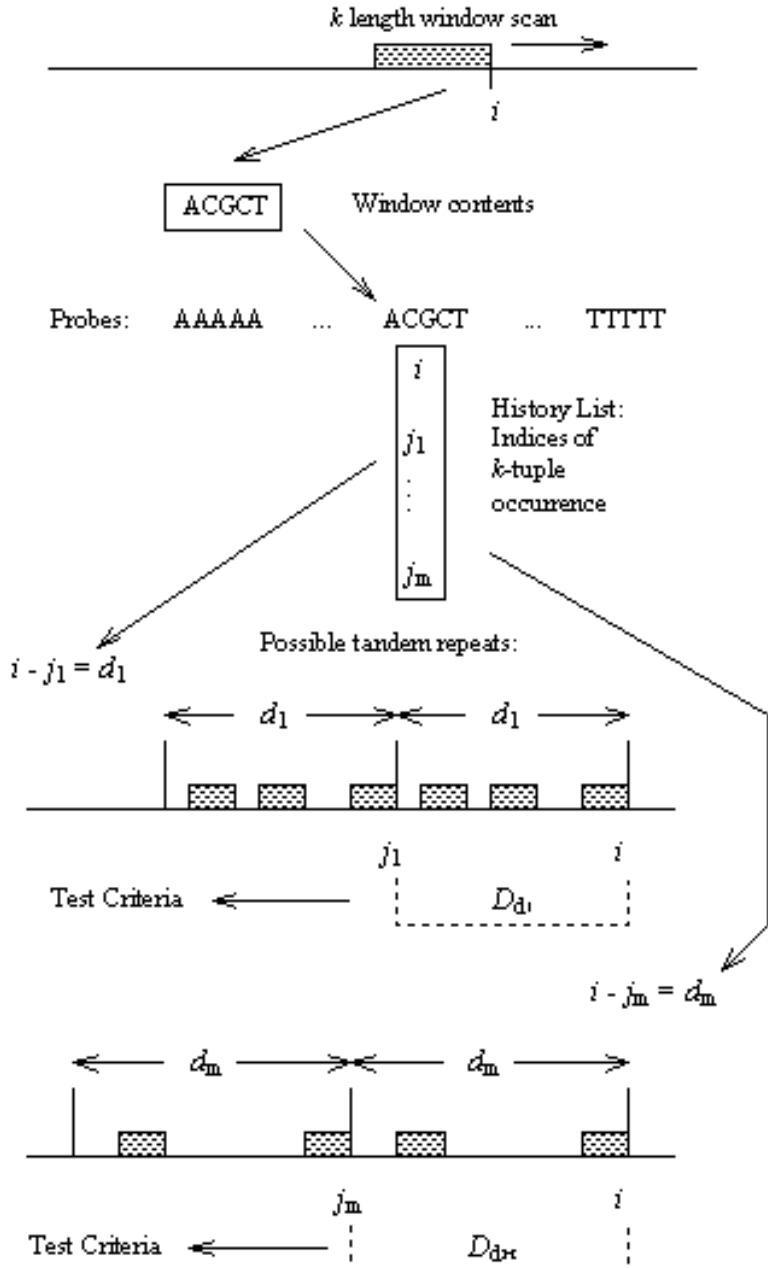
# Short Tandem Repeats not just junk any more!



# **STR identification**

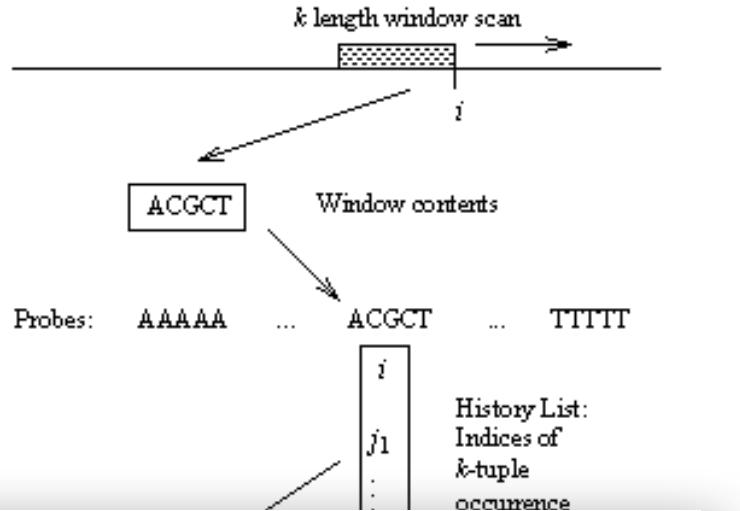
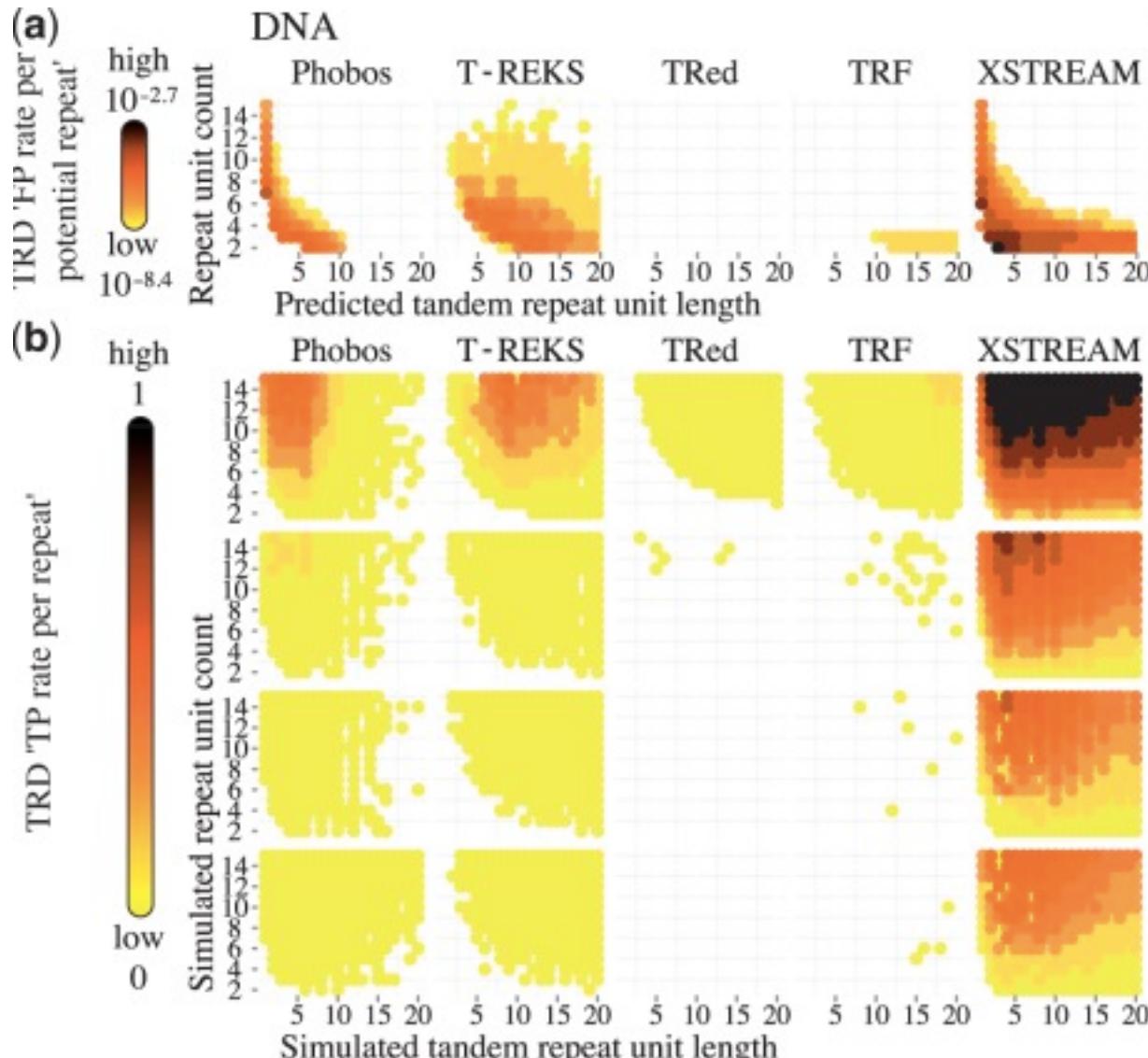
# de novo identification

# **ANDAM REPEATS FINDER**



**Figure 2.** Tandem Repeats are detected by scanning the sequence with a small window, determining the distance between exact matches and testing the statistical criteria.

# de novo identification



## Repeat or not repeat?—Statistical validation of tandem repeat prediction in genomic sequences

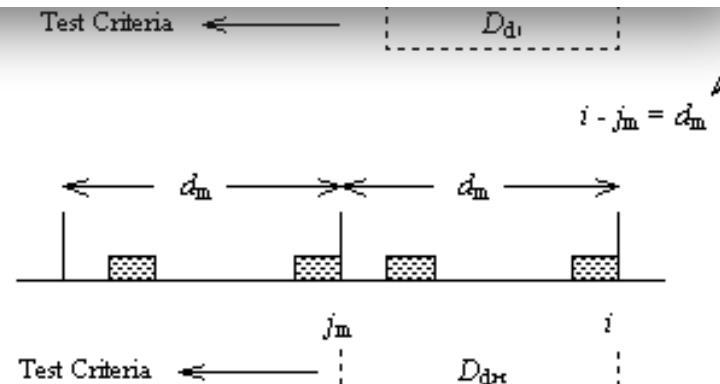
Elke Schaper<sup>1,2,3,\*</sup>, Andrey V. Kajava<sup>4</sup>, Alain Hauser<sup>5</sup> and Maria Anisimova<sup>1,2,\*</sup>

<sup>1</sup>Computer Science Department, ETH Zürich, Universitätstrasse 6, CH-8092 Zürich, Switzerland, <sup>2</sup>Swiss Institute of Bioinformatics, Quartier Sorge—Bâtiment Genopode, CH-1015 Lausanne, Switzerland,

<sup>3</sup>Environmental System Science Department, ETH Zürich, Universitätstrasse 16, CH-8092 Zürich, Switzerland,

<sup>4</sup>Centre de Recherches de Biochimie Macromolculaire, CNRS, University of Montpellier 1 and 2, FR-34293 Montpellier, France and <sup>5</sup>Seminar for Statistics, ETH Zürich, Rämistrasse 101, CH-8092 Zürich, Switzerland

Received May 20, 2012; Revised June 22, 2012; Accepted July 5, 2012



**Figure 2.** Tandem Repeats are detected by scanning the sequence with a small window, determining the distance between exact matches and testing the statistical criteria.

# ovo identification

Sequence analysis

## TRAL: tandem repeat annotation library

Elke Schaper<sup>1,2,3,\*</sup>, Alexander Korsunsky<sup>4</sup>, Jūlija Pečerska<sup>2,3,5</sup>,  
Antonio Messina<sup>6</sup>, Riccardo Murri<sup>6</sup>, Heinz Stockinger<sup>2</sup>, Stefan Zoller<sup>2,3</sup>,  
Ioannis Xenarios<sup>1,2</sup> and Maria Anisimova<sup>2,7</sup>

<sup>1</sup>Vital-IT group, SIB Swiss Institute of Bioinformatics, Quartier Sorge, 1015 Lausanne, Switzerland, <sup>2</sup>SIB Swiss Institute of Bioinformatics, Quartier Sorge, 1015 Lausanne, Switzerland, <sup>3</sup>Department of Computer Science, ETH Zürich, 8092 Zürich, Switzerland, <sup>4</sup>Graz University of Technology, Institute of Molecular Biotechnology, 8010 Graz, Austria, <sup>5</sup>Department of Biosystems Science and Engineering, ETH Zürich, 4058 Basel, Switzerland, <sup>6</sup>Services and Support for Science IT, University of Zürich, 8057 Zürich, Switzerland and <sup>7</sup>Institute of Applied Simulations, School of Life Sciences und Facility Management, Zürich University of Applied Sciences, 8820 Wädenswil, Switzerland

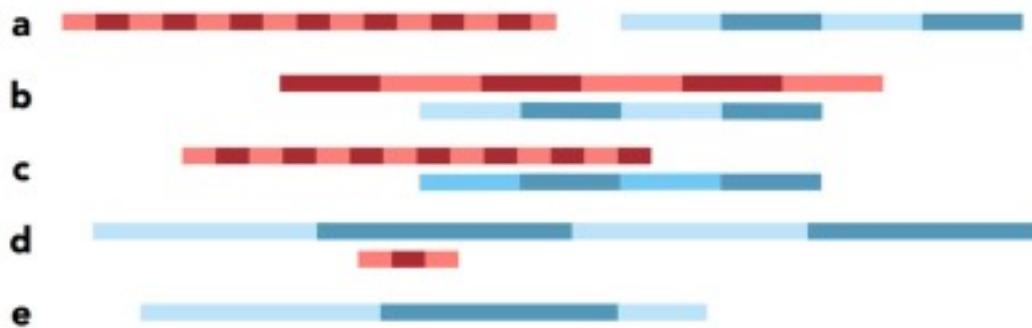
A ...GCTACGTACCTACTAACCTACCTACCTAA...

TR unit alignment

ACGTACCT  
AC-TACCT  
ACCTACCT

C

Sequence



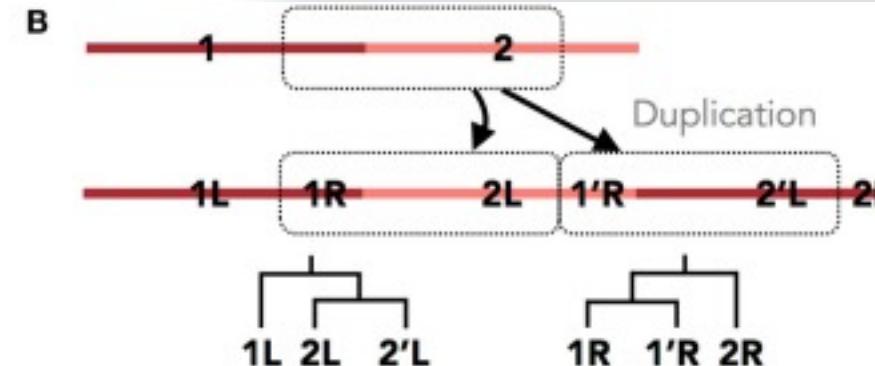
a Non-overlapping TRs

b Coinciding TRs of different coverage

c Overlapping, but not coinciding TRs

d TR in a TR

e Incomplete TR units



## cal validation of genomic sequences

and Maria Anisimova<sup>1,2,\*</sup>

<sup>1</sup>, CH-8092 Zürich, Switzerland, <sup>2</sup>Swiss Institute of Bioinformatics, Quartier Sorge, 1015 Lausanne, Switzerland, <sup>3</sup>Department of Computer Science, ETH Zürich, 8092 Zürich, Switzerland, <sup>4</sup>Graz University of Technology, Institute of Molecular Biotechnology, 8010 Graz, Austria, <sup>5</sup>Department of Biosystems Science and Engineering, ETH Zürich, 4058 Basel, Switzerland, <sup>6</sup>Services and Support for Science IT, University of Zürich, 8057 Zürich, Switzerland and <sup>7</sup>Institute of Applied Simulations, School of Life Sciences und Facility Management, Zürich University of Applied Sciences, 8820 Wädenswil, Switzerland

# STR genotyping

# Challenges in STR genotyping

(a)

True allele    ACTACGATC**ACACACACAC**----TGTGATCGAT

Observed  
(with PCR)  
ACTACGATC**ACACACACAC**----TGTGATCGAT  
ACTACGATC**ACACACACACAC**--TGTGATCGAT  
ACTACGATC**ACACACACACACACT**TGTGATCGAT  
ACTACGATC**ACACACAC**----TGTGATCGAT

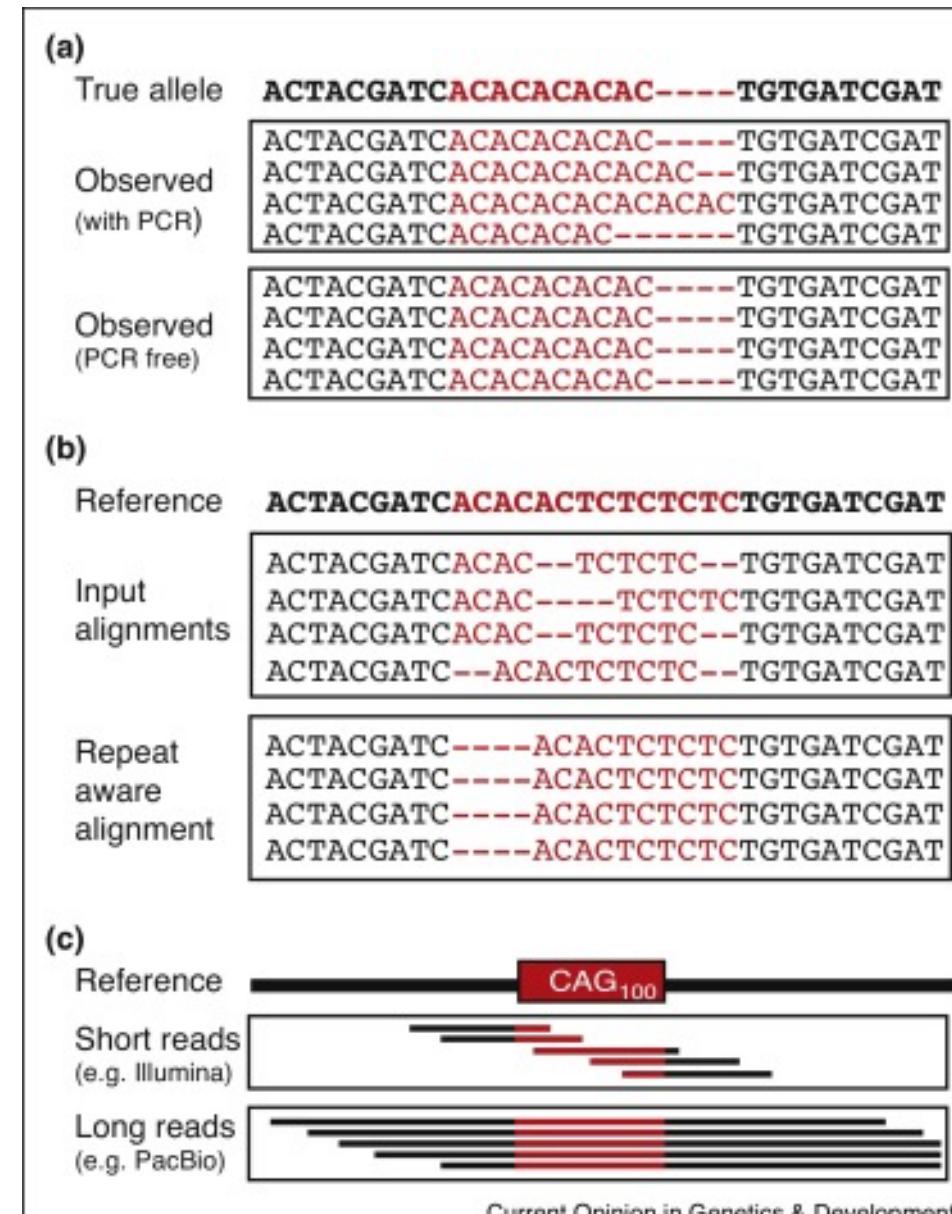
Input  
alignments  
ACTACGATC**ACAC**--TCTCTC--TGTGATCGAT  
ACTACGATC**ACAC**----TCTCTCTGTGATCGAT  
ACTACGATC**ACAC**--TCTCTC--TGTGATCGAT  
ACTACGATC--**ACACTCTCT**--TGTGATCGAT

(c)

Reference    

Short reads  
(e.g. Illumina)  


# Challenges and solutions in STR genotyping



# STR genotyping

Review

Special Issue: Human Genetics

## The overdue promise of short tandem repeat variation for heritability

Maximilian O. Press, Keisha D. Carlson, and Christine Queitsch

Department of Genome Sciences, University of Washington, Foege Building S-250, Box 355065, 3720 15th Avenue NE, Seattle, WA 98195-5065, USA

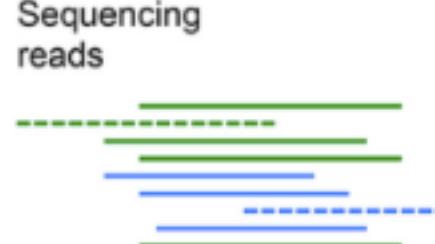
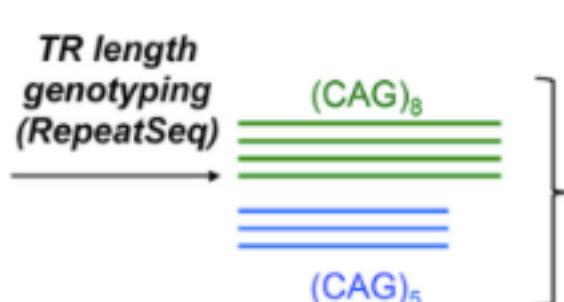
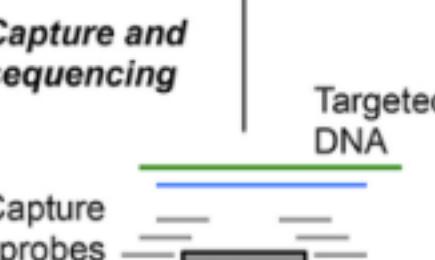
Name	Sequencing reads 	TR length genotyping (RepeatSeq) 	Reported efficiency	Limitations	Ref.
lobSTR			0.2% of reads are informative	Depends on depth of sequencing and length of reads	[37]
RepeatSeq			Not reported	Depends on depth of sequencing and length of reads	[36]
STRViper	Capture and sequencing 	Targeted DNA Gene transcription start site (TSS)	Not reported	Cannot call STR unit-number genotypes	[38]
Array capture	Capture probes 		2.2% informative reads	Low enrichment for STR-spanning reads	[34]
SureSelect RNA probe capture		TR ±1 Kb	27% informative reads	Expensive probe design, captured only 60% of targeted STRs	[35]

Table 1

## Tools for genome-wide profiling of short tandem repeats

	Method description	Language/ platform	STR ref	Multi sample	Models PCR errors	Uses existing alignment
<b>STR-specific tools</b>						
lobSTR [18]	Genotypes 1–6 bp STRs using flank-mapping and maximum likelihood-based genotyping.	C++	Y	Y	Y	Y <sup>a</sup>
RepeatSeq [20]	Genotypes 1–6 bp STRs using existing alignments followed by Bayesian model selection.	C++	Y	N	Y	Y
STR-FM [19]	Genotypes 1–6 bp STRs using flank-mapping and maximum likelihood-based genotyping.	Python/ Galaxy	Y	N	Y	N
TSSV [65]	Targeted profiling of complex allelic variants in pure and mixed genomes	Python	Y <sup>b</sup>	N	N	N
popSTR [23*]	Genotypes 2–6 bp STRs using existing alignments and trains individual-specific error models	C++	Y	Y	Y	Y
HipSTR [24*]	Genotypes 1–6 bp STRs using repeat aware realignment and learns locus-specific error models. Phases STRs onto SNP haplotypes	C++	Y	Y	Y	Y
TRhist [60]	Detects expanded 2–6 bp STRs using hybrid short and long read approach	Java	N	N	N	N
VNTRseek [56]	Detects minisatellites ( $\geq 7$ bp)	C	Y	N	Y	N
<b>STR profiling in cancer</b>						
MSIseq [66]	MSI decision tree classifier using existing genotype calls for tumor samples	R	Y	NA	N	Y <sup>c</sup>
MSIsensor [67]	Identifies changes in length distributions of STRs in paired tumor/normal samples	C++	Y	NA	N	Y
mSINGS [68]	Identifies changes in fraction of unstable STRs for tumor vs. a control population	Python	Y	NA	N	Y

Table 1

## Tools for genome-wide profiling of short tandem repeats

	Method description	Language/ platform	STR ref	Multi sample	Models PCR errors	Uses existing alignment
<b>STR-specific tools</b>						
lobSTR [18]	Genotypes 1–6 bp STRs using flank-mapping and maximum likelihood-based genotyping.	C++	Y	Y	Y	Y <sup>a</sup>
RepeatSeq [20]	Genotypes 1–6 bp STRs using existing alignments followed by Bayesian model selection.	C++	Y	N	Y	Y
STR-FM [19]	Genotypes 1–6 bp STRs using flank-mapping and maximum likelihood-based genotyping.	Python/ Galaxy	Y	N	Y	N
TSSV [65]	Targeted profiling of complex allelic variants in pure and mixed genomes	Python	Y <sup>b</sup>	N	N	N
popSTR [23*]	Genotypes 2–6 bp STRs using existing alignments and trains individual-specific error models	C++	Y	Y	Y	Y
HipSTR [24*]	Genotypes 1–6 bp STRs using repeat aware realignment and learns locus-specific error models. Phases STRs onto SNP	C++	Y	Y	Y	Y

**Method**

TRhis N N

VNTR Y N

STR MSIs N Y<sup>c</sup>

MSIs N Y

mSIN <sup>1</sup>Harvard–MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; <sup>2</sup>Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA; <sup>3</sup>Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv 69978, Israel N Y**lobSTR: A short tandem repeat profiler for personal genomes**Melissa Gymrek,<sup>1,2</sup> David Golan,<sup>2,3</sup> Saharon Rosset,<sup>3</sup> and Yaniv Erlich<sup>2,4</sup><sup>1</sup>Harvard–MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; <sup>2</sup>Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA; <sup>3</sup>Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv 69978, Israel

# ERLICH LAB

*"The Future Ain't What It Used to Be" (Yogi Berra)*

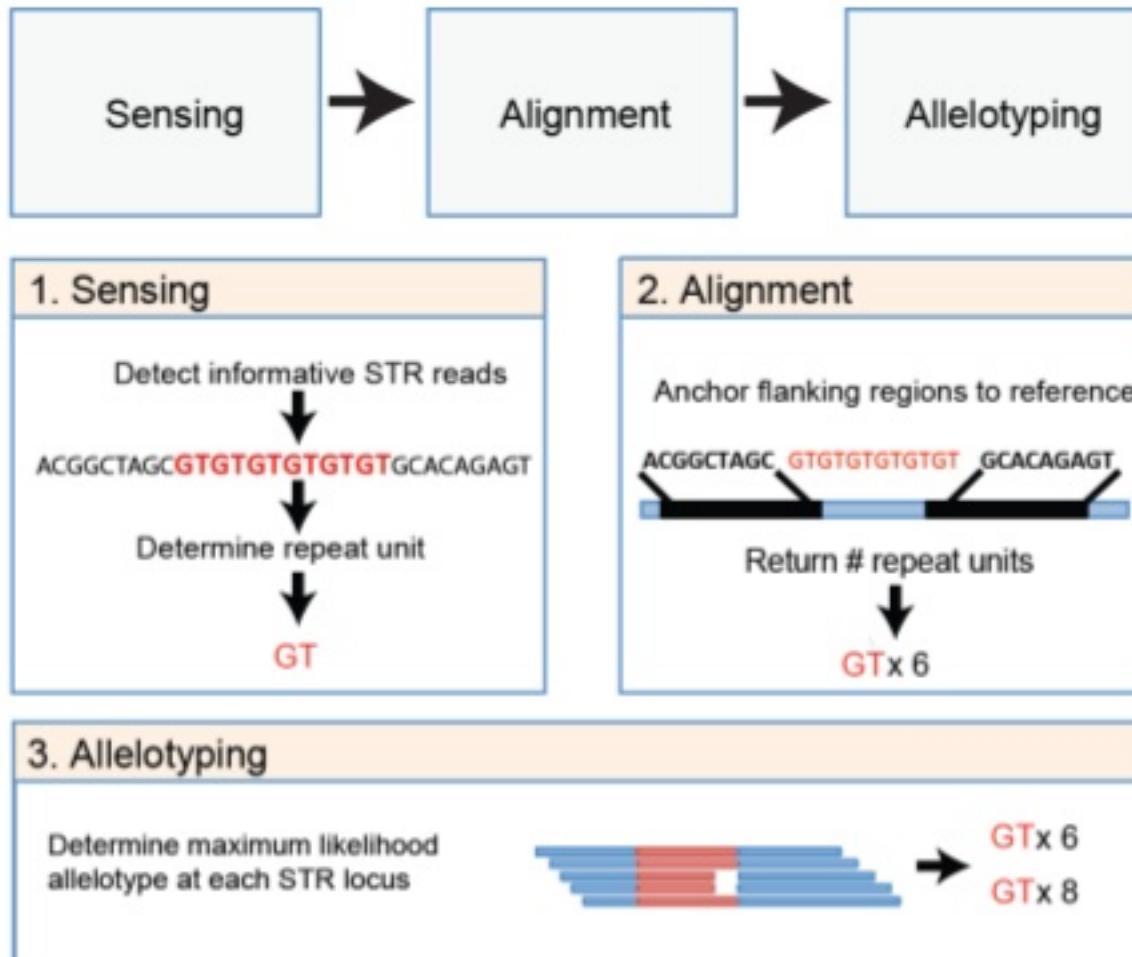


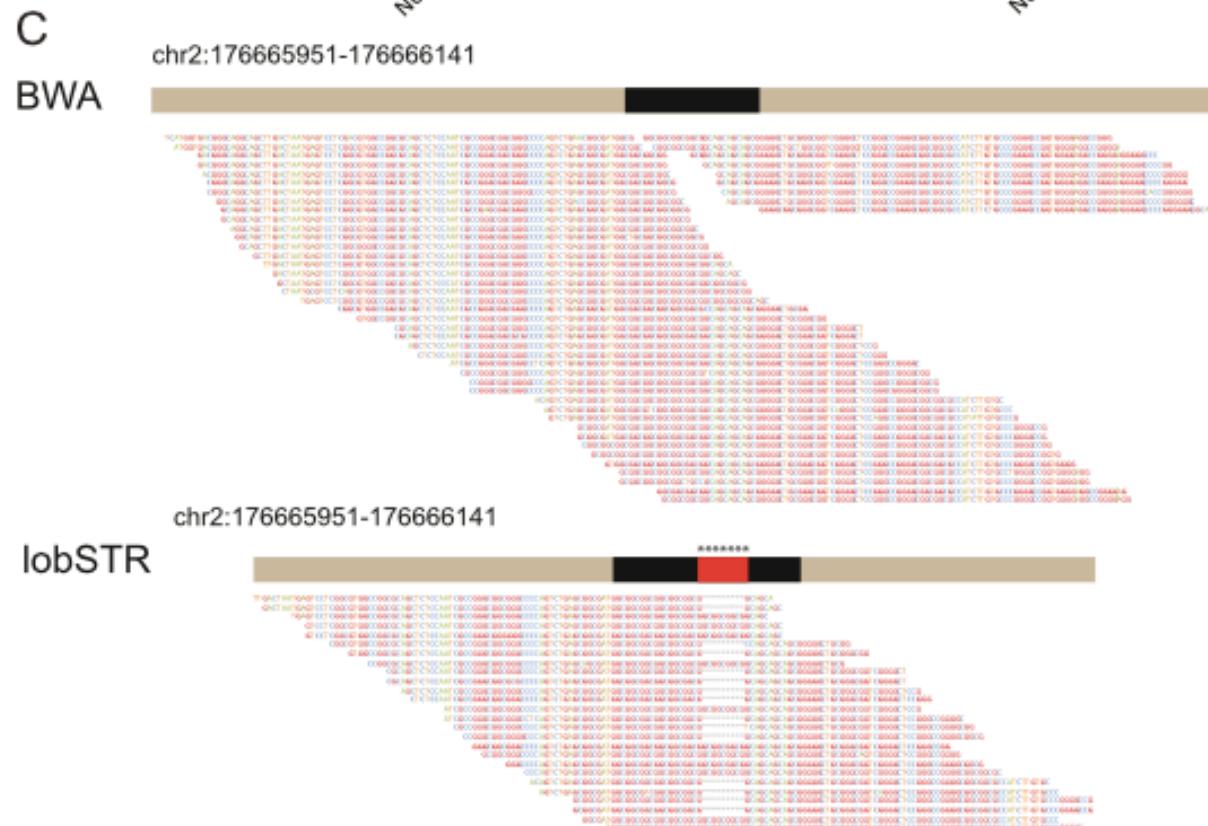
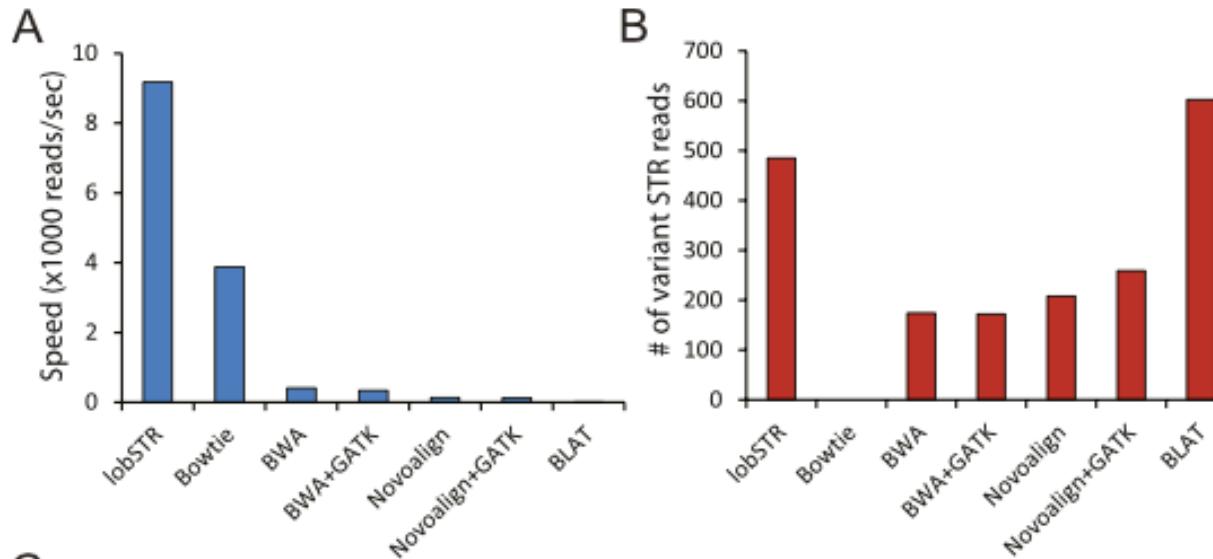
## Welcome to the Gymrek Lab



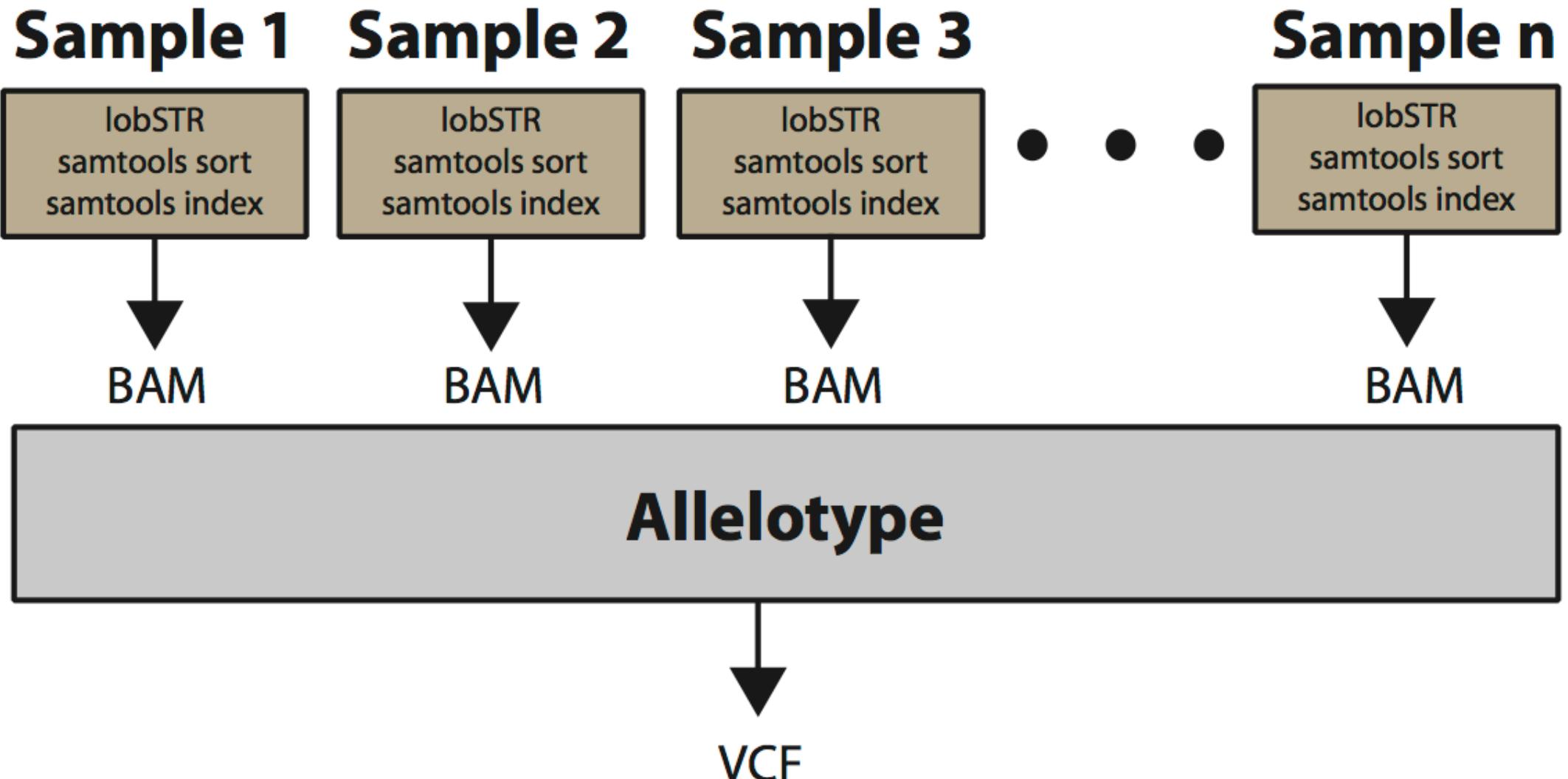
The Gymrek Lab is located at the University of California San Diego. We are part of the Department of Medicine's Division of Genetics and the Department of Computer Science and Engineering. We are also a member of the [Bioinformatics and Systems Biology program](#) at UCSD.

# lobSTR - Whole genome solution for STR genotyping

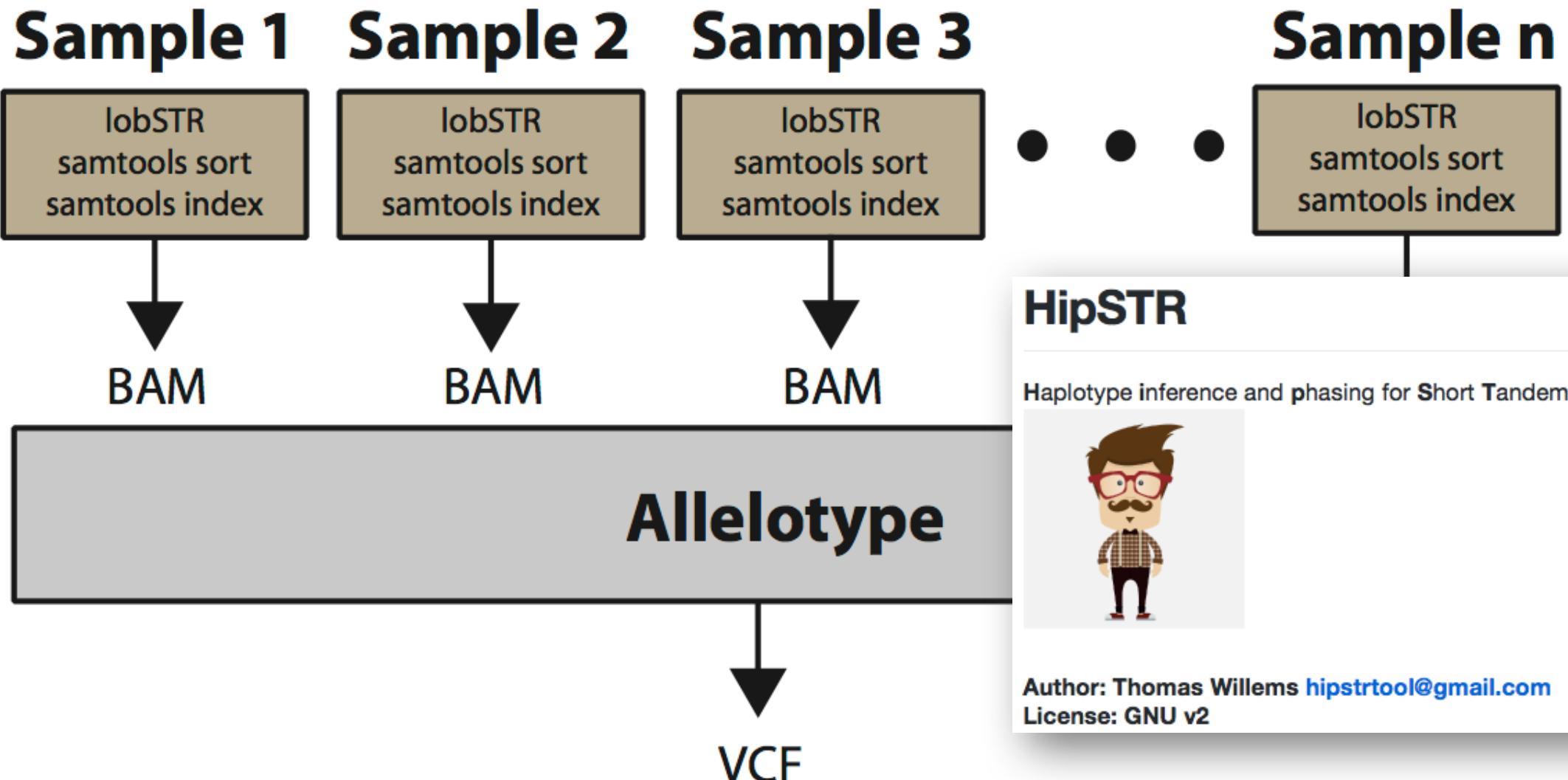




# population-wide STR genotyping



# population-wide STR genotyping



practice

# Bed files

- oncogenes.bed
- tsgs.bed
- strs.bed
- How many different STR alleles have cancer genes?