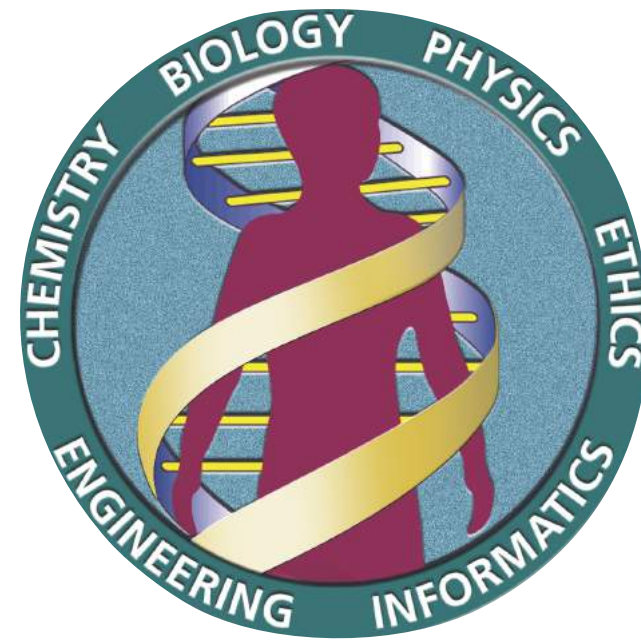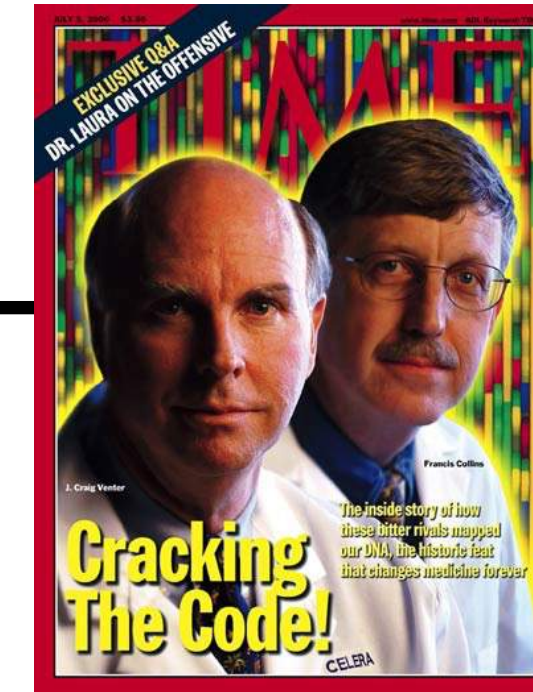# Big Data, Big Science, **Big Impact!**

educator slides

Human Genome Project

1990

2003

- identified the sequence of the ~3 billion chemical bases in a human genome

- mapped the location of ~21,000-23,000 human genes

- predicted intron/exon boundaries for each gene

- in many cases, identified known or predicted amino acid sequence for the corresponding proteins.

HapMap identified the location of ~4 million common human SNPs (single nucleotide polymorphisms) and their frequencies across 4 populations
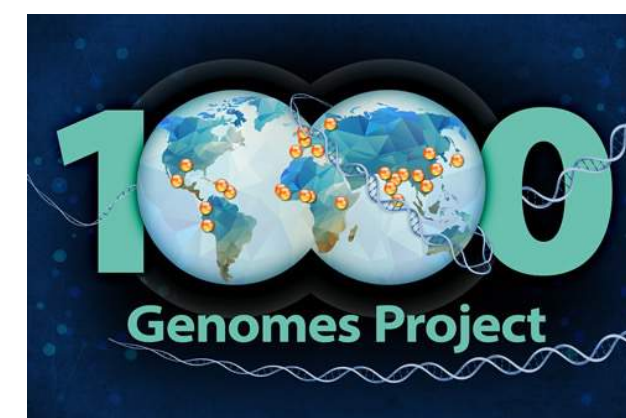
Genome-wide Association Studies (GWAS) use SNPs identified by HapMap to find common genetic variants that affect health and disease

2002   2004   2006   2008   2010   2012   2014   2016

1000 Genomes created a human DNA variation reference at a higher resolution than HapMap.

Identified an almost complete set of DNA variants genome-wide across 26 different populations.
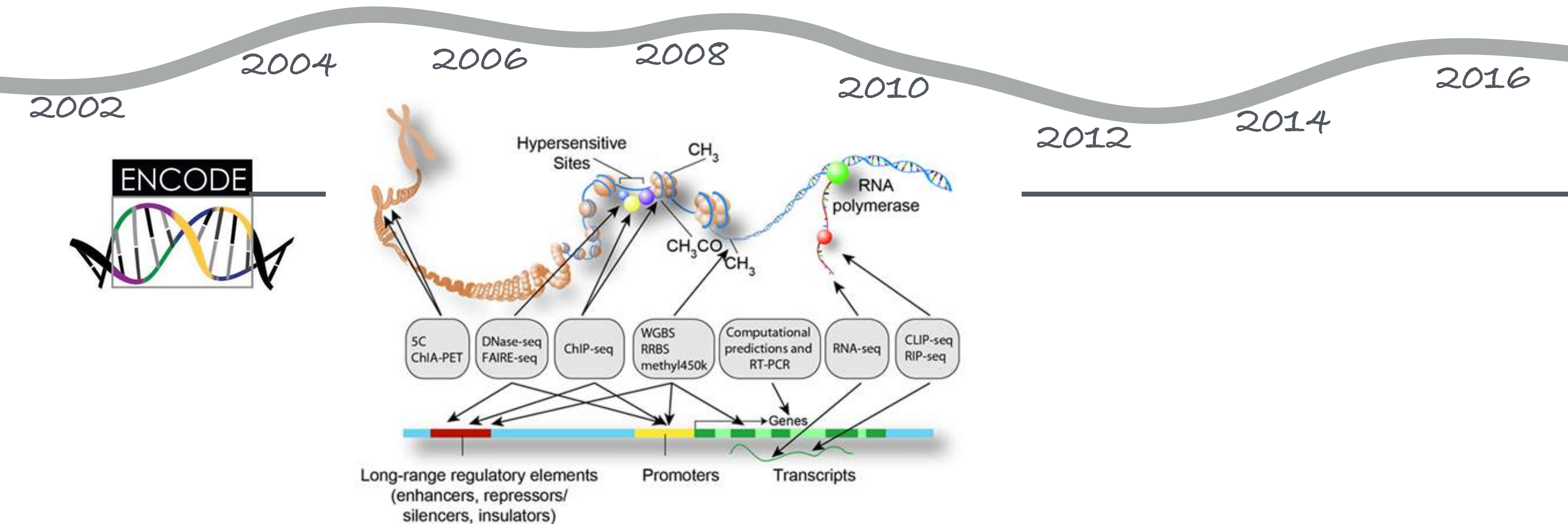
Serves as a reference when analyzing DNA changes identified in individuals with genetic disorders.

HudsonAlpha
INSTITUTE FOR BIOTECHNOLOGY

ENCODE (encyclopedia of DNA elements) hopes to identify all of the functional parts of the genome, determining what sequences regulate the transcriptional activity of the genes.

It builds upon the findings of the Human Genome Project to develop the operating manual for the human genome.

2002 2004 2006 2008 2010 2012 2014 2016

TCGA

2002   2004   2006   2008   2010   2012   2014   2016

TCGA data describes

33 DIFFERENT TUMOR TYPES

...including

10 RARE CANCERS

...based on paired tumor and normal tissue sets collected from

11,000 PATIENTS

...using

7 DIFFERENT DATA TYPES

TCGA (the cancer genome atlas) identified genomic changes (mutations, structural variations, etc.) in over 33 types of human cancer.

Sought to better understand how DNA mutations caused cells to become cancerous.

Worked to determine how that understanding could lead to better prevention, diagnosis and treatment of cancer.

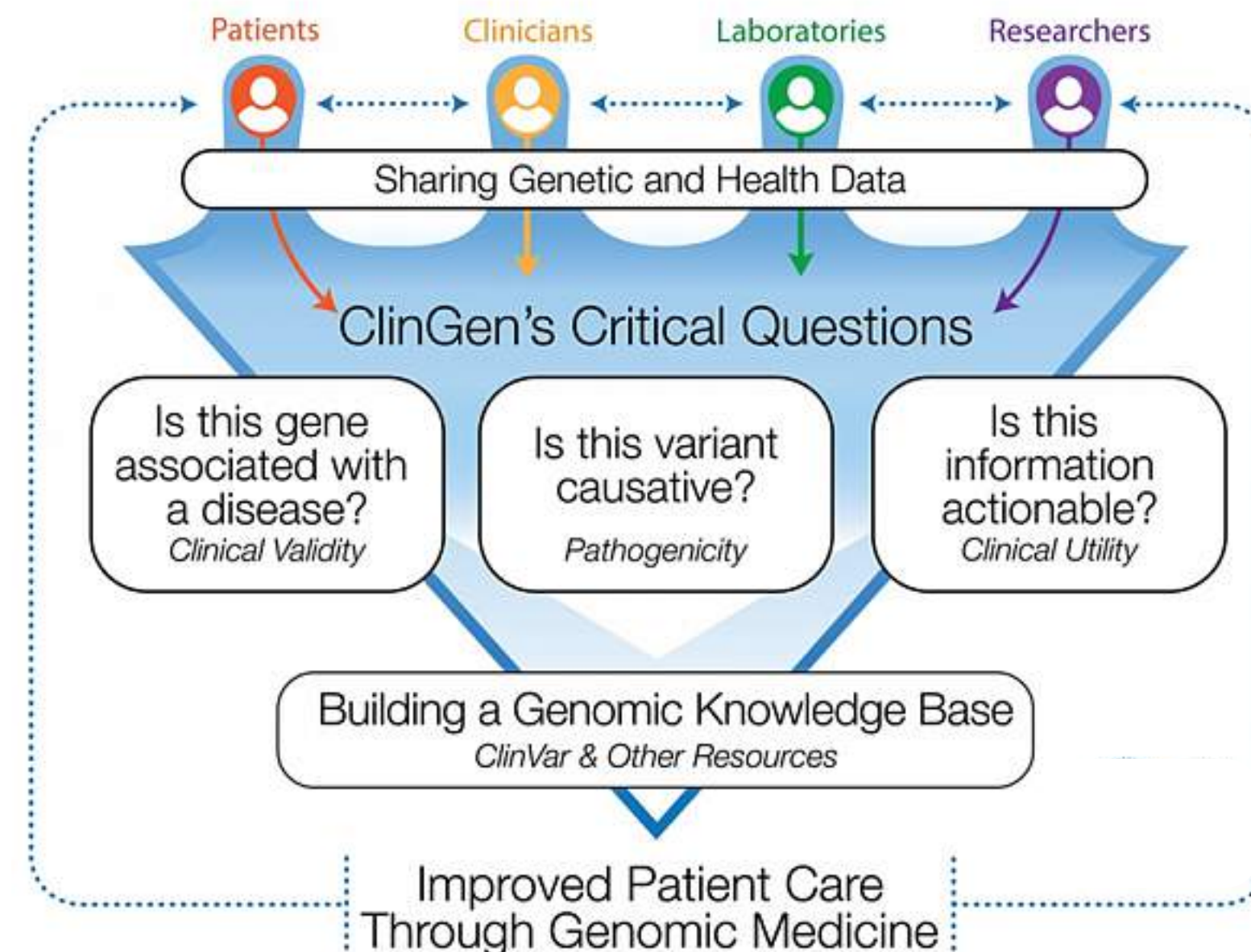**HudsonAlpha**
INSTITUTE FOR BIOTECHNOLOGY

ClinVar is a database where individuals submit human DNA changes and their assessment of its functional and clinical consequence.
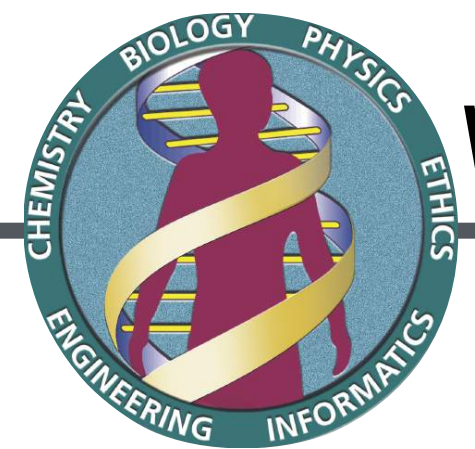

2002   2004   2006   2008   2010   2012   2014   2016

ClinVar data is a key part of a resource called **ClinGen**. It seeks to help scientists and physicians understand the relationship between DNA change and human health to impact patient care.

within a gene?

TCGA

associated with cancer?

commonly occurring?

ClinVar
Clinically relevant variation

CTGATGGTATGGGG
AGGTACGGCTGTCA
AGGGCTGGGATAAA
CATGGTGCATCTGA
CAGGTTGGTATCAA
GCACTGACTCTCTC

clinical interpretation?

2004   2006   2008
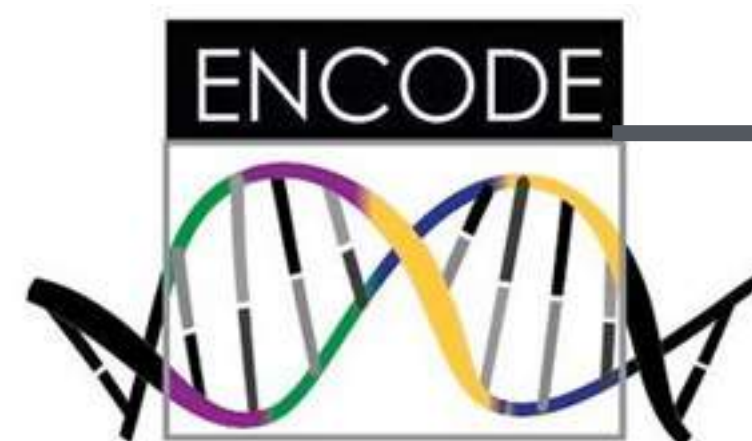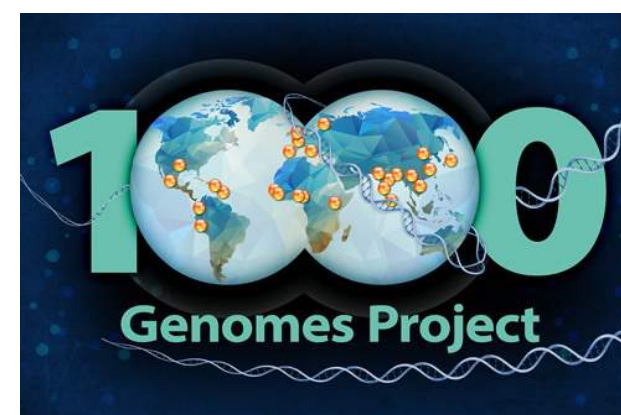
2010

2002   2014

2012

ENCODE

impact transcription and gene activity?

100
Genomes Project

frequency across world populations?

Key Questions answered by each project
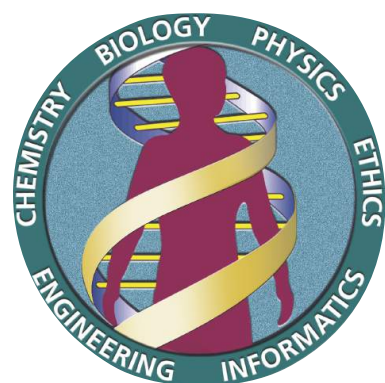
ClinGen
Clinical Genome Resource

clinical use?

What data did these big science projects provide to answer our question linking the DNA change and bitter taste perception?

A DNA change is identified at position 141972905 on chromosome 7. Does this DNA change lead to increased sensitivity to the taste of bitter foods like Brussels sprouts?

TAS2R38 gene

ENCODE identified transcription factors that bind at the TAS2R38 promoter

28 associations

BHLHE40	CHD1	CHD2	CHD7	CTCF	CUX1	EP300	ETS1	EZH2
GTF2F1	H2AFZ	HDAC1	HDAC2	KAT2B	MAX	MAZ	MXI1	PHF8	POLR2A
RAD21	RBBP5	RCOR1	SAP30	SIN3A	SMC3	SUZ12	USF2	ZNF143

A DNA change is identified at position 141972905 on chromosome 7. Does this DNA change lead to increased sensitivity to the taste of bitter foods like Brussels sprouts?
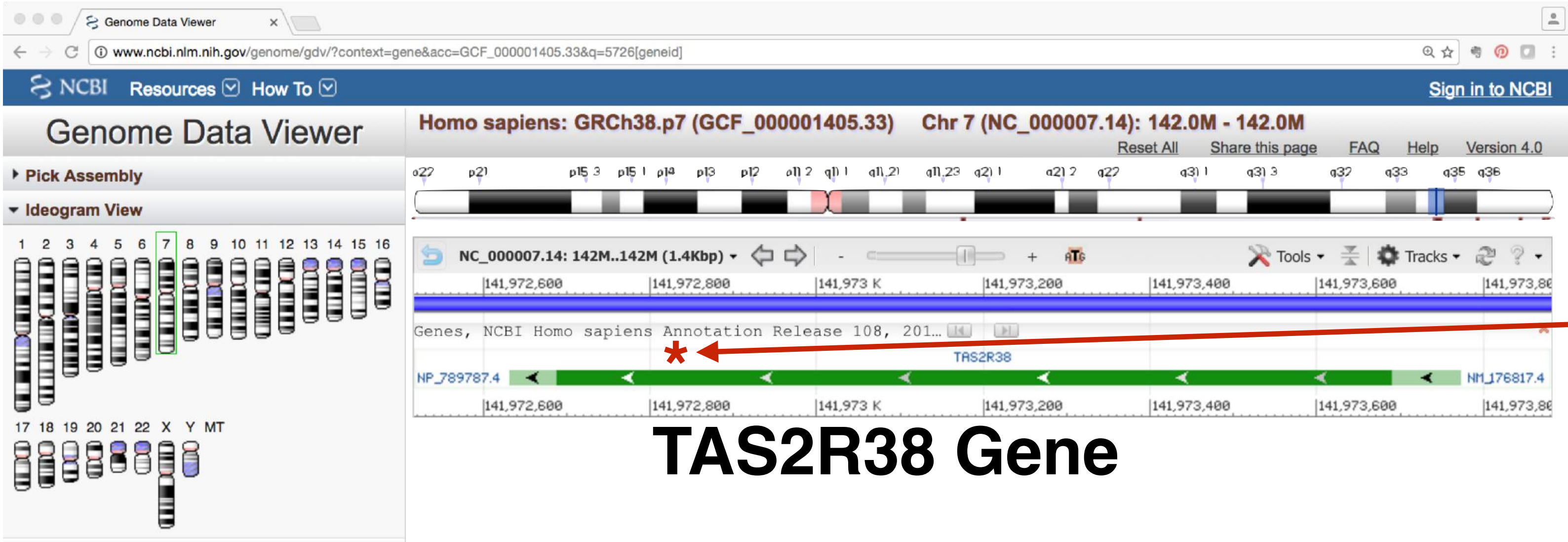
TAS2R38 gene - chromosome 7

141972905
(reference sequence)

141,972,900

141,972,910

| 5' | G | A | T | G | A | A | G | G | C | A | G | C | A | C | A | G | G | A | T | G | A | 3' transcribed strand |

| 3' | C | T | A | C | T | T | C | C | G | T | C | G | T | G | T | C | C | T | A | C | T | 5' nontranscribed strand |

| 3' | C | U | A | C | U | U | C | C | G | U | C | G | U | G | U | C | C | U | A | C | U | 5' RNA |

Ile   Phe   Ala   Ala   Cys   Ser   Ser   amino acid

HudsonAlpha
INSTITUTE FOR BIOTECHNOLOGY
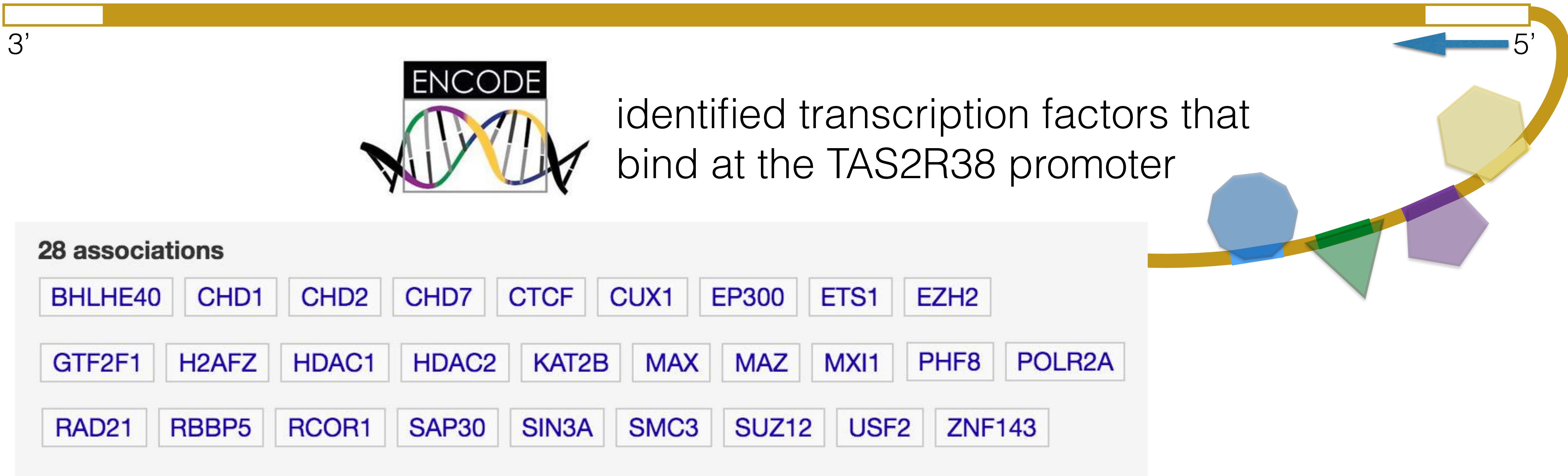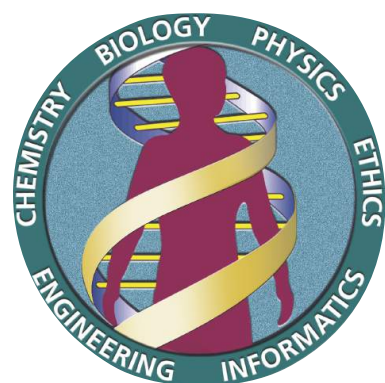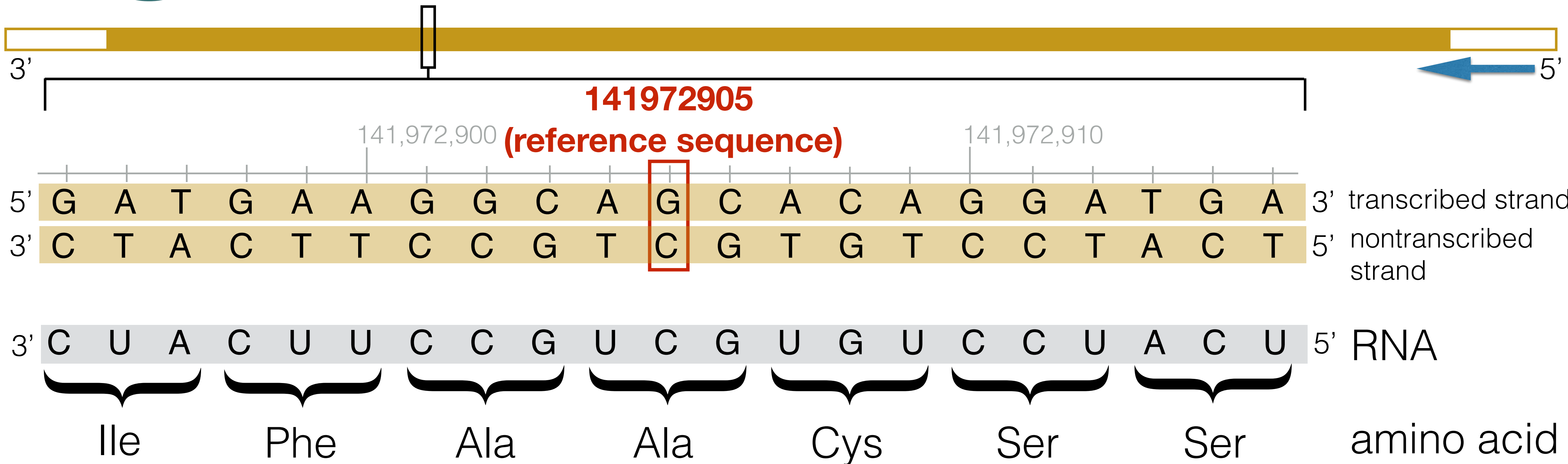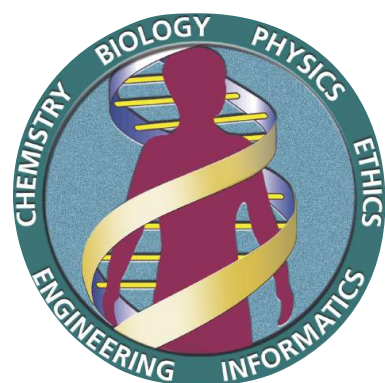
A DNA change is identified at position 141972905 on chromosome 7. Does this DNA change lead to increased sensitivity to the taste of bitter foods like Brussels sprouts?
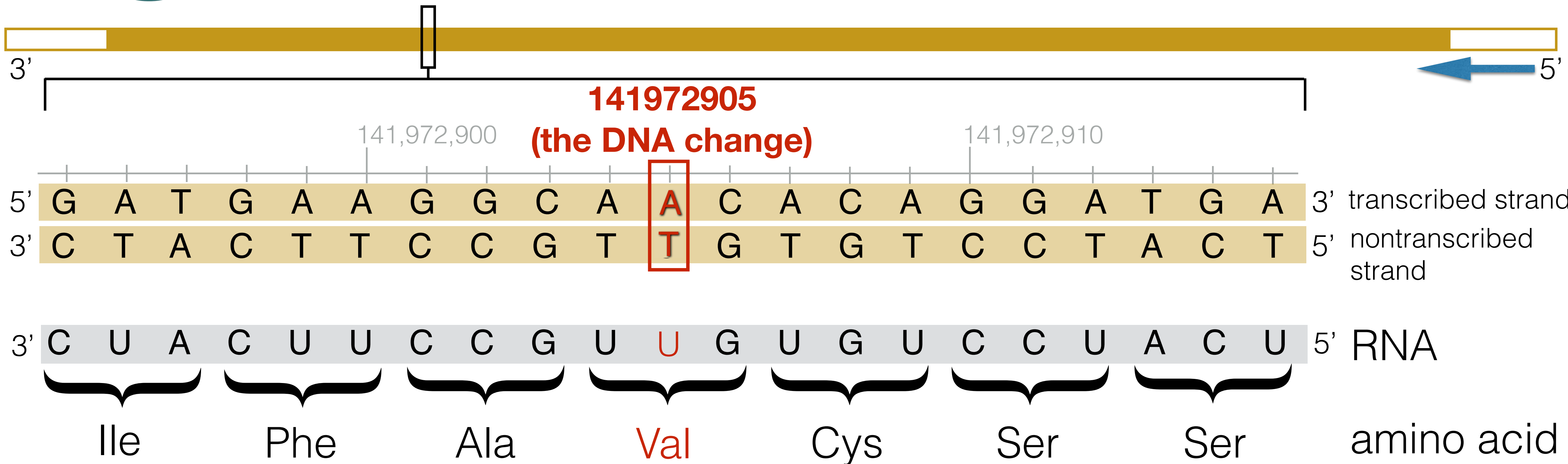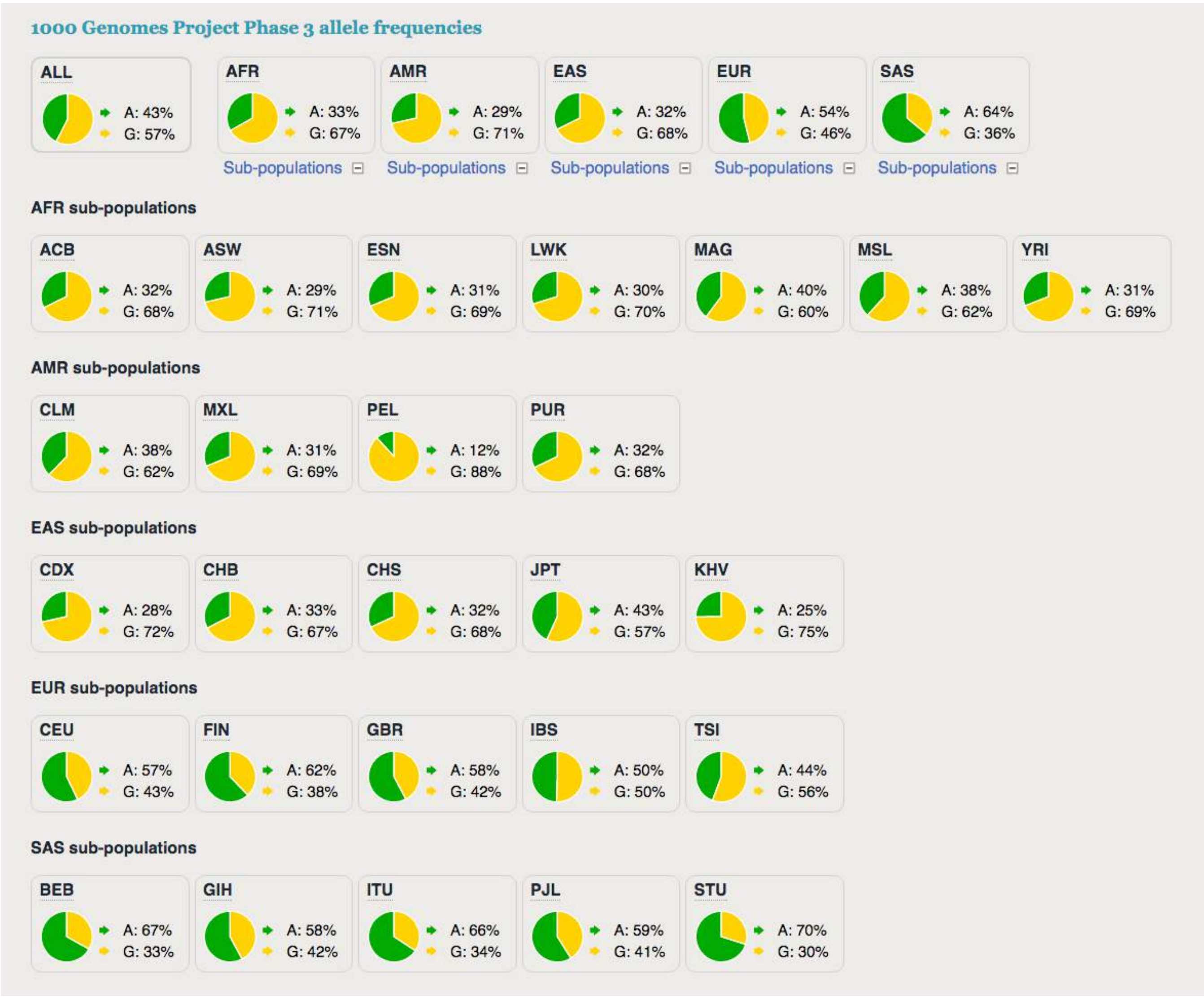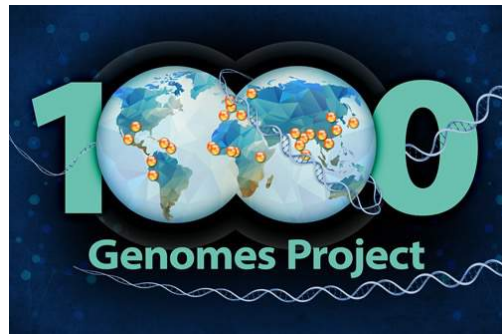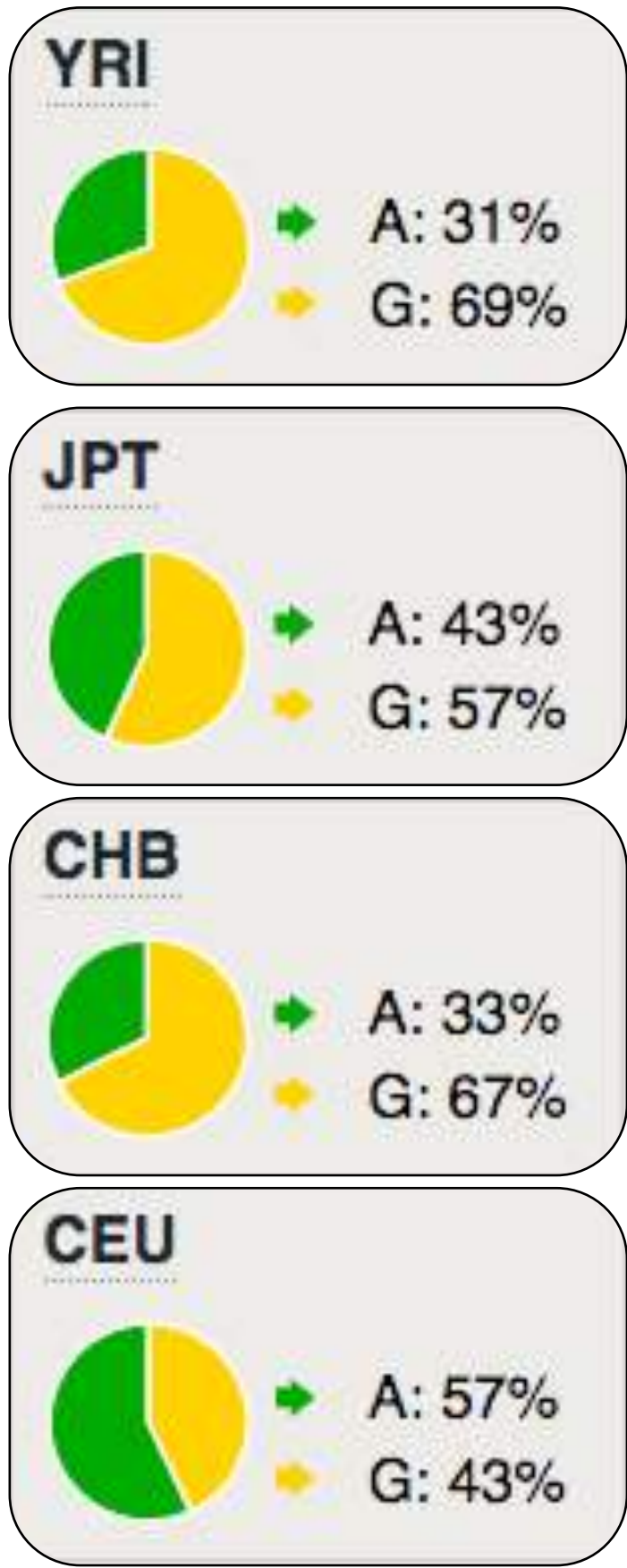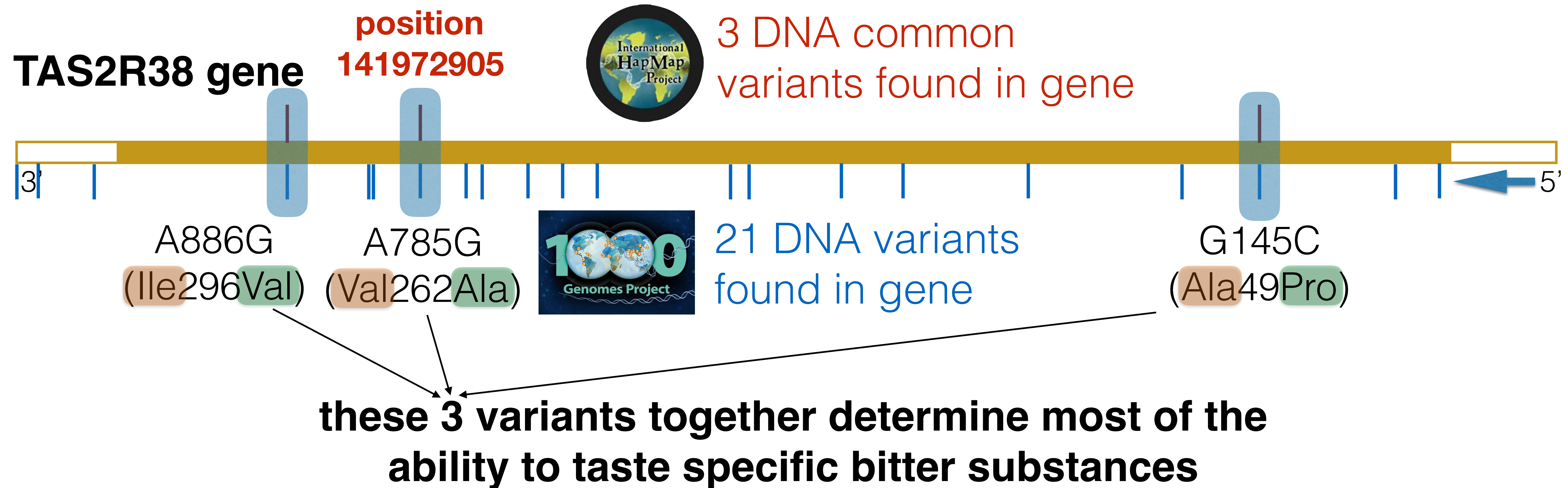
What is the frequency of the DNA change?

International HapMap Project

YRI
- A: 31%
- G: 69%

JPT
- A: 43%
- G: 57%

CHB
- A: 33%
- G: 67%

CEU
- A: 57%
- G: 43%

1000 Genomes Project

1000 Genomes Project Phase 3 allele frequencies

ALL	AFR	AMR	EAS	EUR	SAS
A: 43% G: 57%	A: 33% G: 67%	A: 29% G: 71%	A: 32% G: 68%	A: 54% G: 46%	A: 64% G: 36%
Sub-populations	Sub-populations	Sub-populations	Sub-populations	Sub-populations	Sub-populations

AFR sub-populations

ACB	ASW	ESN	LWK	MAG	MSL	YRI
A: 32% G: 68%	A: 29% G: 71%	A: 31% G: 69%	A: 30% G: 70%	A: 40% G: 60%	A: 38% G: 62%	A: 31% G: 69%

AMR sub-populations

CLM	MXL	PEL	PUR
A: 38% G: 62%	A: 31% G: 69%	A: 12% G: 88%	A: 32% G: 68%

EAS sub-populations

CDX	CHB	CHS	JPT	KHV
A: 28% G: 72%	A: 33% G: 67%	A: 32% G: 68%	A: 43% G: 57%	A: 25% G: 75%

EUR sub-populations

CEU	FIN	GBR	IBS	TSI
A: 57% G: 43%	A: 62% G: 38%	A: 58% G: 42%	A: 50% G: 50%	A: 44% G: 56%

SAS sub-populations

BEB	GIH	ITU	PJL	STU
A: 67% G: 33%	A: 58% G: 42%	A: 66% G: 34%	A: 59% G: 41%	A: 70% G: 30%

HudsonAlpha
INSTITUTE FOR BIOTECHNOLOGY

The tongue is covered with bumps called **papillae**.

Each papilla contains multiple **taste buds**.

Taste buds are filled with **gustatory cells** - the cells that do the tasting. The tip of each gustatory cell protrudes through a pore on the surface of the tongue.

Nerves carry signals from the gustatory cells to the brain.

The tip of each gustatory cell is covered with an assortment of **bitter taste receptors**, which can detect a wide variety of compounds. Stimulation of any of these receptors sends a signal to the brain: bitter!

Kim et al. J. Dent Res (2004)

TAS2R38 protein

- **the DNA change at position 141972905 that substitutes valine for alanine at amino acid 262 changes the shape of the protein's inner pore**

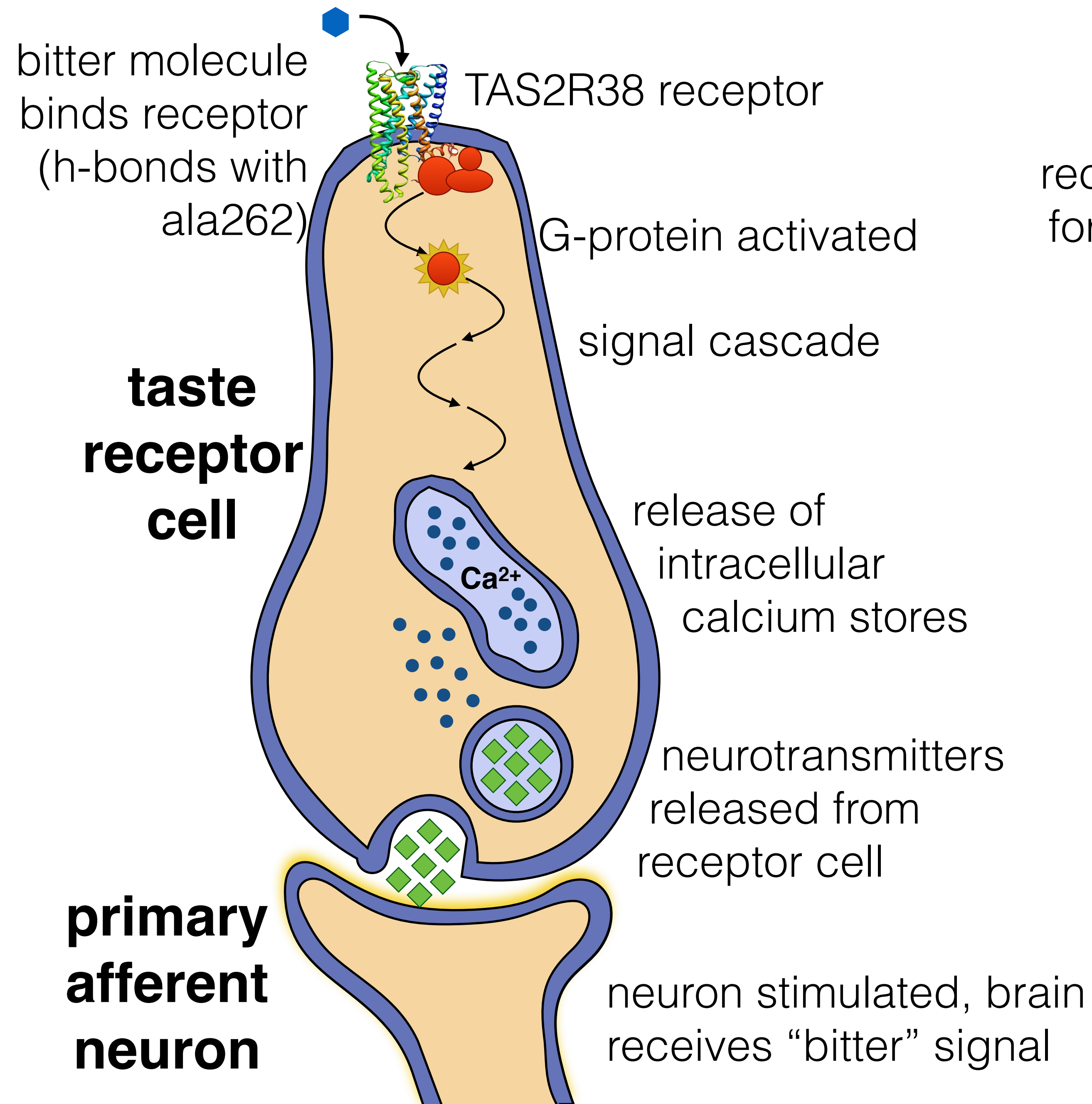http://learn.genetics.utah.edu/content/basics/ptc/images/taste.png

HudsonAlpha
INSTITUTE FOR BIOTECHNOLOGY

TAS2R38 (ala 262)

bitter molecule binds receptor (h-bonds with ala262)

TAS2R38 receptor

taste receptor cell

G-protein activated

signal cascade

release of intracellular calcium stores

Ca^{2+}

neurotransmitters released from receptor cell

primary afferent neuron

neuron stimulated, brain receives "bitter" signal

TAS2R38 (val 262)

bitter molecule unable to bind receptor (no h-bond formed with val262)

TAS2R38 receptor

cascade not initiated "bitter" message not transmitted

Ca^{2+}