BLAST and Finding Data

Sept 9, 2015

ANNOUNCEMENTS

BLAST

Steps in BLAST

BLAST

1. Build Lookup table

Preprocess: Build a *lookup table* of size $|\Sigma|^w$ for all w-length words in D

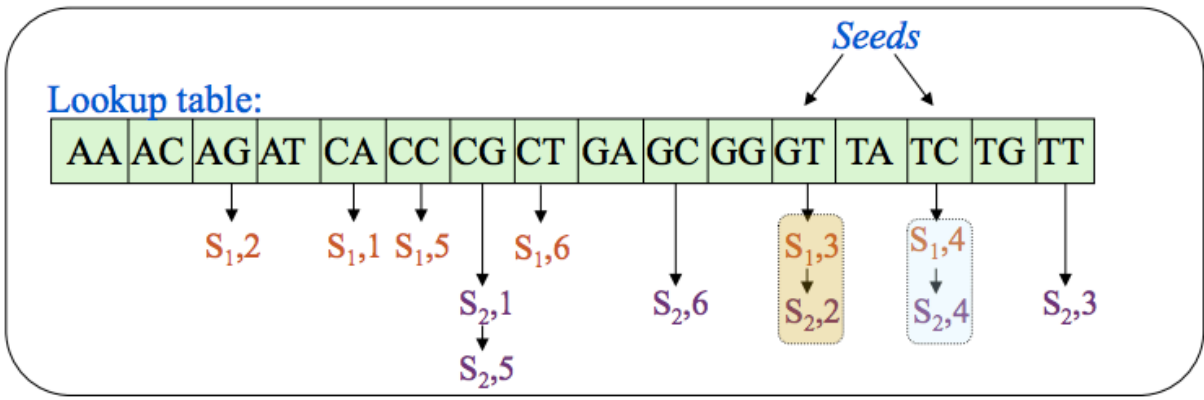$\Sigma=\{A,C,G,T\}$

$w = 2$

➔ 4^2 (=16) entries in lookup table

Lookup table:

AA	AC	AG	AT	CA	CC	CG	CT	GA	GC	GG	GT	TA	TC	TG	TT

BLAST

2. Filter low complexity and Identify Seeds

BLAST

3. Bidirectional extension – (Smith Waterman algorithm)

BLAST

4. Rank and report

BLAST

Stats

$$E = Kmne^{-\lambda S}$$

BLAST

Stats

$$p = 1 - e^{-E}$$

BLAST

Is my p-value significant?

	H_o true	H_o false
Reject H_o	**Type 1 error (false pos)**	**Correct!**
Accept H_o	Correct!	Type 2 error (false neg)

BLAST

Is my p-value significant?

	H_o true	H_o false
Reject H_o	**Type 1 error (false pos)**	**Correct!**
Accept H_o	Correct!	Type 2 error (false neg)

BLAST null: There is no match between query and
database entry

BLAST

Multiple testing correction

Finding Data

Read data
- http://www.ebi.ac.uk/ena
- http://www.ncbi.nlm.nih.gov/sra
- http://metagenomics.anl.gov/?page=MetagenomeSelect

Assembly (and other) Data
- http://useast.ensembl.org/info/data/ftp/index.html
- http://www.ncbi.nlm.nih.gov/genome/
- http://datadryad.org/
- http://figshare.com/

Finding Data

Human Stuff

- http://www.ncbi.nlm.nih.gov/clinvar/
- http://www.ncbi.nlm.nih.gov/omim
- http://snpedia.com/index.php/SNPedia

Show results for all profiles ⌄

Journal	*23andMe White Paper*
Study Size	👥
Replications	None
Contrary Studies	None
Applicable Ethnicities	European
Marker	rs2937573

A study of roughly 80,000 individuals with European ancestry who participated in 23andMe research surveys identified a genetic marker associated with sensitivity to the sound of other people chewing food. The marker rs2937573 is located near a gene (TENM2) that may play a role in the brain. Individuals with the GG genotype at rs2937573 had about 1.2 times higher odds of being sensitive to the sound of chewing, compared to individuals with the AG genotype. Individuals with the AA genotype had about 1.2 times lower odds of being sensitive.

Who	Genotype	Genetic Result
Kate MacManes, Lilly Mendel (Mom)	GG	Slightly higher odds of being sensitive to the sound of chewing.
Lauren MacManes, Owen MacManes, Patrick MacManes	AG	Typical odds of being sensitive to the sound of chewing.
Matthew MacManes, Greg Mendel (Dad)	AA	Slightly lower odds of being sensitive to the sound of chewing.

Sensitivity to the sound of chewing (misophonia)

Journal	23andMe White Paper
Study Size	👥
Replications	None
Contrary Studies	None
Applicable Ethnicities	European
Marker	rs2937573

A study of roughly 80,000 individuals with European ancestry who participated in 23andMe research surveys identified a genetic marker associated with sensitivity to the sound of other people chewing food. The marker rs2937573 is located near a gene (TENM2) that may play a role in the brain. Individuals with the GG genotype at rs2937573 had about 1.2 times higher odds of being sensitive to the sound of chewing, compared to individuals with the AG genotype. Individuals with the AA genotype had about 1.2 times lower odds of being sensitive.

Who	Genotype	Genetic Result
Kate MacManes, Lilly Mendel (Mom)	GG	Slightly higher odds of being sensitive to the sound of chewing.
Lauren MacManes, Owen MacManes, Patrick MacManes	AG	Typical odds of being sensitive to the sound of chewing.
Matthew MacManes, Greg Mendel (Dad)	AA	Slightly lower odds of being sensitive to the sound of chewing.

rs2937573 *[Homo sapiens]*

1.

GCCCAGTCAAAAGTGGCAAGTGCCC[A/G]CACTGTGACTAAGTAAGATGGTGTA

Chromosome:	5:167044193
Gene:	TENM2 (GeneView)
Functional Consequence:	intron variant
Validated:	by 1000G,by 2hit 2allele,by cluster,by frequency,by hapmap,by submitter
Global MAF:	G=0.3990/1998
HGVS:	NC_000005.10:g.167044193G>A, NC_000005.9:g.166471198G>A, XM_005265950.1:c.-189-29049G>A, XM_006714897.1:c.-189-29049G>A, XM_011534604.1:c.-189-29049G>A

Sensitivity to the sound of chewing (misophonia)

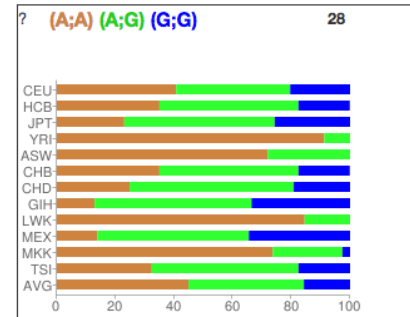| Journal	23andMe White Paper
Study Size	👥👥👥
Replications	None
Contrary Studies	None
Applicable Ethnicities	European
Marker	rs2937573

A study of roughly 80,000 individuals with European ancestry who participated in 23andMe research surveys identified a genetic marker associated with sensitivity to the sound of other people chewing food. The marker rs2937573 is located near a gene (TENM2) that may play a role in the brain. Individuals with the GG genotype at rs2937573 had about 1.2 times higher odds of being sensitive to the sound of chewing, compared to individuals with the AG genotype. Individuals with the AA genotype had about 1.2 times lower odds of being sensitive.

Who	Genotype	Genetic Result
Kate MacManes, Lilly Mendel (Mom)	GG	Slightly higher odds of being sensitive to the sound of chewing.
Lauren MacManes, Owen MacManes, Patrick MacManes	AG	Typical odds of being sensitive to the sound of chewing.
Matthew MacManes, Greg Mendel (Dad)	AA	Slightly lower odds of being sensitive to the sound of chewing.

☐ rs2937573 *[Homo sapiens]*
1.

GCCCAGTCAAAAGTGGCAAGTGCCC[A/G]CACTGTGACTAAGTAAGATGGTGTA

Chromosome:	5:167044193
Gene:	TENM2 (GeneView)
Functional Consequence:	intron variant
Validated:	by 1000G,by 2hit 2allele,by cluster,by frequency,by hapmap,by submitter
Global MAF:	G=0.3990/1998
HGVS:	NC_000005.10:g.167044193G>A, NC_000005.9:g.166471198G>A, XM_005265950.1:c.-189-29049G>A, XM_006714897.1:c.-189-29049G>A, XM_011534604.1:c.-189-29049G>A

Sensitivity to the sound of chewing (misophonia)

Journal	23andMe White Paper
Study Size	👥👥👥
Replications	None
Contrary Studies	None
Applicable Ethnicities	European
Marker	rs2937573

A study of roughly 80,000 individuals with European ancestry who participated in 23andMe research surveys identified a genetic marker associated with sensitivity to the sound of other people chewing food. The marker rs2937573 is located near a gene (TENM2) that may play a role in the brain. Individuals with the GG genotype at rs2937573 had about 1.2 times higher odds of being sensitive to the sound of chewing, compared to individuals with the AG genotype. Individuals with the AA genotype had about 1.2 times lower odds of being sensitive.

Who	Genotype	Genetic Result
Kate MacManes, Lilly Mendel (Mom)	GG	Slightly higher odds of being sensitive to the sound of chewing.
Lauren MacManes, Owen MacManes, Patrick MacManes	AG	Typical odds of being sensitive to the sound of chewing.
Matthew MacManes, Greg Mendel (Dad)	AA	Slightly lower odds of being sensitive to the sound of chewing.

rs2937573 *[Homo sapiens]*

1.

GCCCAGTCAAAAGTGGCAAGTGCCC[A/G]CACTGTGACTAAGTAAGATGGTGTA

Chromosome:	5:167044193
Gene:	TENM2 (GeneView)
Functional Consequence:	intron variant
Validated:	by 1000G,by 2hit 2allele,by cluster,by
Global MAF:	G=0.3990/1998
HGVS:	NC_000005.10:g.167044193G>A, NC_00
	XM_005265950.1:c.-189-29049G>A, XM_
	XM_011534604.1:c.-189-29049G>A

ALFRED

The ALlele FREquency Database
ALFRED is a resource of gene frequency data on human populations supported by the U. S. National Science Foundation.

→ Home → Ethics → Search → Summaries → Documentation → Register → Contact Us

Polymorphism Information

Name	ALFRED UID	Locus Name	Locus Symbol
rs2937573	SI368946K	rs2937573 is intergenic between RPLP0P9 and ODZ2	rs2937573

Fst	Avg Het	# Populations Typed
0.181	0.407	51

Synonyms: *rs2937573* ;
Frequency on Map: GoogleMap Help

Frequency Display Formats: Graph Table
Estimated Heterozygosity: Graph
Frequency Download: Tab Delimited ◆ Arlequin Help
External Resources: dbSNP rs# Record PharmGKB Variant Information Record
References: See References
Polymorphism Description: This is a A/G SNP
Alleles:

Allele Name	Allele Symbol	Description
A	A	5' - gtcaaaagtggcaagtgccc *A* cactgtgactaagtaagatg - 3'
G	G	5' - gtcaaaagtggcaagtgccc *G* cactgtgactaagtaagatg - 3'

References:
- Kenneth K. Kidd et al. "Data unpublished".