

SNP Resources: Finding SNPs Databases and Data Extraction

Mark J. Rieder, PhD
Robert J. Livingston, PhD

NIEHS Variation Workshop
January 30-31, 2005

Genotype - Phenotype Studies

Typical Approach:

"I have candidate gene/region and samples ready to study. Tell me what SNPs to genotype."

Other questions:

How do I know I have *all* the SNPs?
What is the validation/quality of the SNPs that are known?
Are these SNPs informative in my population/sample?

What do I need to know for selecting the "best" SNPs?
How do I pick the "best" SNPs?

What information do I need to characterize a SNP for genotyping?

Minimal SNP information for genotyping/characterization

- What is the SNP? Flanking sequence and alleles.
 - ✓ FASTA format
 - >snp_name
 - ACGGAGTAGCCAG
 - [A/G]
 - ACTGGGATAGAAC
- dbSNP reference SNP # (rs #)
- Where is the SNP mapped? Exon, promoter, UTR, etc
 - ✓ picture of gene with mapped to the gene structure.
- How was it discovered? Method
- What assurances do you have that it is real? Validated how?
- What population – African, European, etc?
- What is the allele frequency of each SNP? Common (>10%), rare
- Are other SNPs associated - redundant? Genotyping data!

Finding SNPs: Databases and Extraction

How do I find and download SNP data for analysis/genotyping?

- Entrez Gene
 - dbSNP
 - Entrez SNP
- HapMap Genome Browser
- NIEHS Environmental Genome Project (EGP) Candidate gene website
- NIEHS web applications and other tools
 - GeneSNPs, PolyDoms, TraFac, PolyPhen, ECR Browser, GVS

NCBI - Database Resource

The screenshot shows the NCBI homepage with a search bar containing 'NOS2A'. Below the search bar, there are several sections: 'What does NCBI do?', 'Hot Spots', 'Influenza Virus Resource', and 'Entrez Gene'. The 'Entrez Gene' section is highlighted, showing a brief description of the resource and a link to search for information centered on the concept of a gene.

Finding SNPs: Where do I start?

http://www.ncbi.nlm.nih.gov/gquery

The screenshot shows the NCBI Entrez search engine results for 'NOS2A'. The search results are displayed in a table with columns for 'Database', 'Accession Number', and 'Description'. The 'Entrez Gene' database is highlighted, showing the gene structure and associated information for NOS2A.

Finding SNPs: Where do I start?

NCBI - Entrez Gene

Entrez Gene
 Search Gene
 Home About POC
 Search Tools Gene Expression Database Cross-species (FTP)
 Mapping Lists Gene Profiles Feedback
 Help Us Contact Us
 Related Sites BLAST Entrez History Genealogy Projects Genealogy Browser GEO

Search Gene
 Home About POC
 Search Tools Gene Expression Database Cross-species (FTP)
 Mapping Lists Gene Profiles Feedback
 Help Us Contact Us
 Related Sites BLAST Entrez History Genealogy Projects Genealogy Browser GEO

Gene: **NOX2A**
 Official Symbol: **NOX2A**, and Names: nitric oxide synthase 2A (inducible, hepatoctes) [Homo sapiens]
 Other Aliases: HEP-NO2, NO2S, NO2S2
 Other Designations: NOS, nsp-12, nitric oxide synthase 2A, nitric oxide synthase, cytochrome P450, cytochrome P-17, Loxasein, [Tq1] 2-q12
 GeneID: 40162

rs100012

Finding SNPs: Entrez Gene

Entrez Gene
 Search Gene
 Home About POC
 Search Tools Gene Expression Database Cross-species (FTP)
 Mapping Lists Gene Profiles Feedback
 Help Us Contact Us
 Related Sites BLAST Entrez History Genealogy Projects Genealogy Browser GEO

Gene: **NOX2A**
 Official Symbol: **NOX2A**, and Names: nitric oxide synthase 2A (inducible, hepatoctes) [Homo sapiens]
 Other Aliases: HEP-NO2, NO2S, NO2S2
 Other Designations: NOS, nsp-12, nitric oxide synthase 2A, nitric oxide synthase, cytochrome P450, cytochrome P-17, Loxasein, [Tq1] 2-q12
 GeneID: 40162

rs100012

dbSNP Geneview

Gene Model (mRNA alignment) information from genome sequence
 Total gene model (contig mRNA transcript): 4
 Contig mma protein mma orientation transcript snp list
 NT_010729.NM_000625.NP_000616 reverse minus strand currently shown
 NT_088665.NM_000625.NP_000616 reverse minus strand
 NT_010729.NM_153292.NP_695024 reverse minus strand
 NT_088665.NM_153292.NP_695024 reverse minus strand

Views | in gene region | cSNP | has frequency | double hit | haplotype tagged

gene model (contig mRNA transcript): NT_010729.NM_000625.NP_000616 reverse minus strand 20, coding

Contig	dbSNP raw position	cluster id	Heterozygosity	Validation	MD	OMIM	Function	dbSNP allele	Protein residue	Codon position	Amino acid position
NT_010729.NM_000625.NP_000616	821320	r7523111	N/D				synonymous	A	Ala [A]	3	1136
NT_010729.NM_000625.NP_000616	821320	r7523111	N/D				coding reference	G	Ala [A]	3	1136

dbSNP Geneview

Gene Model (mRNA alignment) information from genome sequence
 Total gene model (contig mRNA transcript): 4
 Contig mma protein mma orientation transcript snp list
 NT_010729.NM_000625.NP_000616 reverse minus strand currently shown
 NT_088665.NM_000625.NP_000616 reverse minus strand
 NT_010729.NM_153292.NP_695024 reverse minus strand
 NT_088665.NM_153292.NP_695024 reverse minus strand

Views | in gene region | cSNP | has frequency | double hit | haplotype tagged

gene model (contig mRNA transcript): NT_010729.NM_000625.NP_000616 reverse minus strand 294, all

Contig	dbSNP raw position	Heterozygosity	Validation	MD	OMIM	Function	dbSNP allele	Protein residue	Codon position	Amino acid position	
NT_010729.NM_000625.NP_000616	820851	r23944210	0.014				untranslated	C/T			
NT_010729.NM_000625.NP_000616	820939	r23944218	0.015				untranslated	A/G			

Finding SNPs: dbSNP validation

Contig	dbSNP raw position	cluster id	Heterozygosity	Validation	MD	OMIM	Function	dbSNP allele	Protein residue	Codon position	Amino acid position
NT_010729.NM_000625.NP_000616	821320	r7523111	N/D				synonymous	A	Ala [A]	3	1136
NT_010729.NM_000625.NP_000616	821320	r7523111	N/D				coding reference	G	Ala [A]	3	1136

Validation status description
 Validated by multiple, independent submissions to the dbSNP Query
 Validated by frequency or genotype data: minor alleles observed in at least two chromosomes.
 Validated by submitter confirmation
 All alleles have been observed in at least two chromosomes.
 HapMap Verified

Contig	dbSNP raw position	cluster id	Heterozygosity	Validation	MD	OMIM	Function	dbSNP allele	Protein residue	Codon position	Amino acid position
NT_010729.NM_000625.NP_000616	820851	r23944210	0.014				untranslated	C/T			
NT_010729.NM_000625.NP_000616	820939	r23944218	0.015				untranslated	A/G			

Finding SNPs: dbSNP database

FASTA sequence (Legend)

gdb|rs3834298|allelePop=25&min=51|taxid=9606|allele=C|O|job=Genomic

GGAGAGGACG CCGCTTTACA GAGGAGAGGG GCGCTAGGGG GACCGACAGCA GTAAATTTGA
 GAGCTTAGG TCGATGTCG TCGTCTGGG TTAGAGAGCG TGTGATGTCG ACTGTGATCG
 TCTGCTCAG ACACATTCAT AGCTCAACT GAKATTATTT TGAAGCTACA CTATGTCTCA
 GGATGATCG APTAAATTA TGAAGCAAAA ATTTTTACA TTTACTATGG GCTGACAAAG
 CAGAGAGA AAGAG

Submitted Frequency for rs3834298

Population ID	Panel	Sample Size (n)	Major Allele Freq	Minor Allele Freq	Estimated Heterozygosity	Genotype Freq	Substituted Heterozygosity	Submission Batch	Submitter
CEP	AD-FANEL	28	C=0.877	C=0.145				NCBI-EGP_AD-051605	EGP-SNP-C
CEP	ASIAN-PANEL	44	C=0.818	C=0.192				NCBI-EGP_ASIAN-051605	EGP-SNP-C
CEP	YORUB-PANEL	24	C=0.975	C=0.125				NCBI-EGP_YORUB-051605	EGP-SNP-C
CEP	CEPH-PANEL	44	C=0.955	C=0.045				NCBI-EGP_CEPH-051605	EGP-SNP-C

Entrez SNP - dbSNP genotype retrieval

NCBI Entrez Gene

Search Gene: [rs3834298]

Gene: **NOXA** *nitric oxide synthase 2A (inducible, hepatoctyes) [Homo sapiens]*

Summary

Official Symbol: **NOXA** and Name: *nitric oxide synthase 2A (inducible, hepatoctyes)* provided by **HGNC**

Gene type: protein coding

RefSeq status: Reviewed

Organism: *Homo sapiens*

Language: *Subtypica; Mexica; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontata; Placentalia; Hominidae; Homo*

Gene aliases: *NO3, NOS, NOS2, HEP-NOS*

Summary: *NOXA* is a reactive free radical which acts as a biologic mediator in several processes, including neuronal apoptosis. This gene encodes a nitric oxide synthase which is expressed in liver and is inducible by a nitric oxide synthase. Three related paralogs are located within the South-Magnon syndrome region on chromosome 17. All two transcripts encode overlapping alleles.

Genomic regions, transcripts, and products

Gene strand: **NC_009047**

Map: [Genomic map showing exons and introns]

Finding SNPs - Gene Genotype Report

NCBI Single Nucleotide Polymorphism

Single Nucleotide Polymorphism

Gene: **NOXA**

Gene ID: **51119**

Gene Symbol: **NOXA**

Chromosome: **17**

Report: **rs3834298**

Linkage Disequilibrium (LD) with other SNPs in this region

Indicates SNPs with **r^2 > 0.8**

Indicates SNPs with **r^2 > 0.6**

Indicates SNPs with **r^2 > 0.4**

Indicates SNPs with **r^2 > 0.2**

Indicates SNPs with **r^2 > 0.0**

Indicates SNPs with **r^2 < 0.0**

Indicates SNPs with **r^2 < -0.2**

Indicates SNPs with **r^2 < -0.4**

Indicates SNPs with **r^2 < -0.6**

Indicates SNPs with **r^2 < -0.8**

Finding SNPs - Gene Genotype Report

dbSNP Genotype and Allele Frequency Report **Build125**

GENE GENOTYPE REPORT

Organism	Gene ID	Gene Symbol	Chr	Report
<i>Homo sapiens</i>	51119	NOXA	17	rs3834298

Genotypes on Human chromosome Subjects

- *PERLEGEN-ADP_EBP_PANEL Pop Desc
- *PERLEGEN-ADP_AFR_PANEL Pop Desc
- *PERLEGEN-ADP_CBN_PANEL Pop Desc
- *CSHL-HAPMAP-HapMap-CEU Pop Desc
- *CSHL-HAPMAP-HapMap-HCB Pop Desc
- *CSHL-HAPMAP-HapMap-JPT Pop Desc
- *CSHL-HAPMAP-HapMap-YRI Pop Desc
- *EGP_SNPS-EGP_YORUB-PANEL Pop Desc
- *EGP_SNPS-EGP_HISL-PANEL Pop Desc
- *EGP_SNPS-EGP_CEPH-PANEL Pop Desc
- *EGP_SNPS-EGP_AD-FANEL Pop Desc
- *EGP_SNPS-EGP_ASIAN-PANEL Pop Desc
- * SNP Detail

Finding SNPs - Gene Genotype Report

dbSNP Genotype and Allele Frequency Report **Build125**

GENE GENOTYPE REPORT

Organism	Gene ID	Gene Symbol	Chr	Report
<i>Homo sapiens</i>	51119	NOXA	17	rs3834298

Genotypes on Human chromosome Subjects

- *PERLEGEN-ADP_EBP_PANEL Pop Desc
- *PERLEGEN-ADP_AFR_PANEL Pop Desc
- *PERLEGEN-ADP_CBN_PANEL Pop Desc
- *CSHL-HAPMAP-HapMap-CEU Pop Desc
- *CSHL-HAPMAP-HapMap-HCB Pop Desc
- *CSHL-HAPMAP-HapMap-JPT Pop Desc
- *CSHL-HAPMAP-HapMap-YRI Pop Desc
- *EGP_SNPS-EGP_YORUB-PANEL Pop Desc
- *EGP_SNPS-EGP_HISL-PANEL Pop Desc
- *EGP_SNPS-EGP_CEPH-PANEL Pop Desc
- *EGP_SNPS-EGP_AD-FANEL Pop Desc
- *EGP_SNPS-EGP_ASIAN-PANEL Pop Desc
- * SNP Detail

Substituted Group: **EGP-Perlegen Individual Sex Match Panel**

Sample	Sex	Group	Genotype	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	Allele 6	Allele 7	Allele 8	Allele 9	Allele 10
NA-043000	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043001	M	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043002	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043003	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043004	M	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043005	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043006	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043007	M	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043008	M	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043009	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043010	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043011	M	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043012	M	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043013	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043014	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043015	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043016	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043017	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043018	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043019	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G
NA-043020	F	EGP	T/T	C	A	G	A	G	G	G	G	G	G

Chrom. gene-coordinates web, or r^2 associated with another substitution

Minimal SNP information for genotyping/characterization

- What is the SNP? Flanking sequence and alleles.
 - ✓ FASTA format
 - >snp_name
 - ACCGAGTAGCCAG
 - [A/G]
 - ACTGGGATAGAAC
 - dbSNP - data is there**
- dbSNP reference SNP # (rs #)
- Where is the SNP mapped? Exon, promoter, UTR, etc
 - ✓ picture of gene with mapped to the gene structure.
- How was it discovered? Method
- What assurances do you have that it is real? Validated how?
- What population – African, European, etc?
- What is the allele frequency of each SNP? Common (>10%), rare
- Are other SNPs associated - redundant? Genotyping data!

Entrez Gene Entry - Entrez SNP

The screenshot shows the NCBI Entrez Gene entry for NOS2A. The 'SNP' section is highlighted with a red box, showing a list of SNPs with their IDs, positions, and frequencies. The 'Genomic regions, transcripts, and products' section is also visible at the bottom.

Entrez SNP - direct dbSNP querying

The screenshot shows the NCBI Entrez SNP interface. A search for 'NOS2A' has been performed, resulting in a list of SNPs. The 'Limits' section is highlighted in a red box, showing options for 'Search all fields' and 'Search all fields, leave the following boxes unchecked (Limits help)'. The 'Function class' section is also highlighted, showing options for 'Coding nonsynonymous', 'reference', 'exception', 'intron', 'coding synonymous', 'locus region', 'mRNA utr', and 'splice site'.

Entrez SNP - Parseable Multi-SNP reports

The screenshot shows the NCBI Entrez SNP interface. A search for 'NOS2A' has been performed, resulting in a list of SNPs. The 'Limits' section is highlighted in a red box, showing options for 'Search all fields' and 'Search all fields, leave the following boxes unchecked (Limits help)'. The 'Function class' section is also highlighted, showing options for 'Coding nonsynonymous', 'reference', 'exception', 'intron', 'coding synonymous', 'locus region', 'mRNA utr', and 'splice site'.

Entrez SNP - Parseable Multi-SNP reports

The screenshot shows the NCBI Entrez SNP interface. A search for 'NOS2A' has been performed, resulting in a list of SNPs. The 'Limits' section is highlighted in a red box, showing options for 'Search all fields' and 'Search all fields, leave the following boxes unchecked (Limits help)'. The 'Function class' section is also highlighted, showing options for 'Coding nonsynonymous', 'reference', 'exception', 'intron', 'coding synonymous', 'locus region', 'mRNA utr', and 'splice site'.

Entrez SNP - Search Limiting Capabilities

The screenshot shows the NCBI Entrez SNP interface. A search for 'NOS2A' has been performed, resulting in a list of SNPs. The 'Limits' section is highlighted in a red box, showing options for 'Search all fields' and 'Search all fields, leave the following boxes unchecked (Limits help)'. The 'Function class' section is also highlighted, showing options for 'Coding nonsynonymous', 'reference', 'exception', 'intron', 'coding synonymous', 'locus region', 'mRNA utr', and 'splice site'.

Entrez SNP - Search Limits

Entrez SNP - Search Limiting Capabilities

Entrez SNP - Query Term Capabilities

Example	Description
BRCA[Gene Name]	Search SNPs on all genes with names starting with the letter 'BRCA' (ie. BRCA1 and BRCA2)
1[SNP]	Search SNPs on all genes with names starting with the letter '1' (ie. BRCA1 and BRCA2)
coding nonsynon[FUNC] AND [CHR]	Search SNPs with function class 'coding nonsynon' located on chromosome 1
[CHR] OR [CHR]	Search all SNPs on chromosome 1 or 2
[CHR] OR [CHR] NOT unknown[METHOD]	Search all SNPs on chromosome 1 or 2 detected by all methods except 'unknown'
[CHR] AND [CHR] OR [CHR] NOT unknown[METHOD]	Search all SNPs with weight 1 on chromosome 1 or 2 detected by all methods except 'unknown' or 'compou'

Entrez SNP - Search Terms Fields

Search Field	Qualifier	Type	Description
Allele	[ALLELE][VARIATION], [VARI]	IUPAC	(Observed alleles) Example: [A][T][C][G]
Chromosome	[CHR]	Text	Map to chromosome number [Available values: 1-22, X, Y, and Un (unknown)] Example: [2][CH] or [X][CHR]
Base Position	[CHRPOS][BPOS]	Integer	Mapped chromosome position; use in conjunction with chromosome field [CHR] Example: [2][CHR] AND [8556398-8556398][CHRPOS]
Create Build ID	[CREATE_BUILD][CBID]	Integer	SNP create build ID Example: [1][CBID]
Publication Date	[CREATEDATE][CDAT][PDAT], [DATE]	Date*	SNP create/publication date Use the format YYYYMMDD; month and day are optional. Example: [2002][CDATE]
Function Class	[FN_CLASS][FUNC]	Text	Function Class: locus region coding nonsynon coding synon reference exception Example: 'coding synon'[FUNC]
Gene Name	[GENE], [GENE_SYMBOL]	Text	Locus link symbol Example: [LPL][GENE]
Protein Description	[PDSC]	Text	Protein description Example: 'fibrinogen alpha 2'[PDSC]
Gene	[GDS]	Text	Gene description Example: 'fibrinogen alpha 2'[GDS]
GO Terms	[GO]	Text	Gene Ontology (GO) terms Example: 'inhibits' [GO term]
Genotype	[GENOTYPE], [GTYPE]	Boolean	Genotype Example: [rs2299417]

Entrez SNP - Search Terms Fields

Submitter Handle	[HANDLE]	Text	Submitter handle Example: 'LEUC[HANDLE]
Accession Key	[ACCESS_KEY][KEY]	Integer	Accession key Example: [918384]
Organism	[ORGANISM][TAX_ID]	Text	Organism name or Taxonomy ID number; use the prefix 'taxid' before the Taxonomy ID number Example: 'human'[ORGANISM], 'mouse'[ORGANISM], or 'taxid5996'[ORGANISM]
Update Build ID	[UPD_BUILD][UBID]	Integer*	SNP update build ID Example: [102][UBID]
Modification Date	[MOD_DATE][UDAT], [MODDATE]	Date*	SNP modification/update date Use the format YYYYMMDD; month and day are optional. Example: [2002][UDATE]
Validation	[VALIDATION]	Text	Validation status: by cluster by submitter by frequency no info Example: 'by cluster'[VALIDATION]

More advanced queries:

2[CHR] AND "coding nonsynon"[FUNC]

ENTREZ **SNP**

Search: SNP [ZICR1 AND coding nonsynon[FUNC]]

dbSNP BUILD 124

Items 1 - 20 of 3208

1: rs17874984 [Homo sapiens]
 chr1:17874984(T)GTTGTTCTGCTATTT(A)CAGTTCACATGAGTGGTTCCTTTTGA

2: rs17874973 [Homo sapiens]
 chr1:17874973(T)GTTGTTCTGCTATTT(A)CAGTTCACATGAGTGGTTCCTTTTGA

3: rs17869125 [Homo sapiens]
 chr1:17869125(T)GTTGTTCTGCTATTT(A)CAGTTCACATGAGTGGTTCCTTTTGA

4: rs17869158 [Homo sapiens]
 chr1:17869158(T)GTTGTTCTGCTATTT(A)CAGTTCACATGAGTGGTTCCTTTTGA

Entrez SNP - Search Terms Fields

Submitter Handle	[HANDLE]	Text	Submitter handle Example: 'LEE:[HANDLE]
Success Rate	[SUCCESS_RATE],[SRATE]	Integer	Success rate(s) Example: '25:SRATE'
Organism	[ORGN],[TAX_ID]	Text	Organism name or Taxonomy ID number; use the prefix 'taxid' before the Taxonomy ID number Example: 'Homo:[ORGN]' , 'mouse:[ORGN]' , or 'tax4906:[ORGN]
Update Build ID	[UPD_BUILD],[UBID]	Integer*	SNP update build ID Example: '102U[UBID]
Modification Date	[UDATE],[UDAT],[MODDATE]	Date*	SNP modification update date Use the format YYYY/MM/DD; month and day are optional. Example: '2002/01 [UDATE]
Validation	[VALIDATION]	Text	Validation status: by cluster by submitter by frequency no info Example: 'by cluster[VALIDATION]

More advanced queries:

2[CHR] AND "coding nonsynon"[FUNC] AND "EGP_SNPS"[HANDLE]

Note: Can also use wildcard (*) characters, AND, OR, and NOT operators

Entrez SNP - Advanced Queries

ENTREZ **SNP**

Search: SNP [2[CHR] AND coding nonsynon[FUNC] AND EGP_SNPS[HANDLE]]

dbSNP BUILD 124

Items 1 - 20 of 66

1: rs28915683 [Homo sapiens]
 chr9:28915683(T)GTTGTTCTGCTATTT(A)CAGTTCACATGAGTGGTTCCTTTTGA

2: rs28915687 [Homo sapiens]
 chr9:28915687(T)GTTGTTCTGCTATTT(A)CAGTTCACATGAGTGGTTCCTTTTGA

3: rs28910528 [Homo sapiens]
 chr9:28910528(T)GTTGTTCTGCTATTT(A)CAGTTCACATGAGTGGTTCCTTTTGA

4: rs28910527 [Homo sapiens]
 chr9:28910527(T)GTTGTTCTGCTATTT(A)CAGTTCACATGAGTGGTTCCTTTTGA

Minimal SNP information for genotyping/characterization

- What is the SNP? Flanking sequence and alleles.
 - FASTA format

```
>snp_name
ACCGAGTAGCCAG      EntrezSNP - better!
[A/G]
ACTGGGATAGAAC
```
- dbSNP reference SNP # (rs #)
- Where is the SNP mapped? Exon, promoter, UTR, etc
 - picture of gene with mapped to the gene structure.
- How was it discovered? Method
- What assurances do you have that it is real? Validated how?
- What population – African, European, etc?
- What is the allele frequency of each SNP? Common (>10%), rare
- Are other SNPs associated - redundant? Genotyping data!

- ### Finding SNPs - Entrez SNP Summary
- dbSNP is useful for investigating detailed information on a small number SNPs - and its good for a picture of the gene
 - Entrez SNP is a direct, fast, database for querying SNP data.
 - Data from Entrez SNP can be retrieved in batches for many SNPs
 - Entrez SNP data can be "limited" to specific subsets of SNPs and formatted in plain text for easy parsing and manipulation
 - More detailed queries can be formed using specific "field tags" for retrieving SNP data

- ### Finding SNPs: Databases and Extraction
- How do I find and download SNP data for analysis/genotyping?
- Entrez Gene
 - dbSNP
 - Entrez SNP
 - HapMap Genome Browser
 - NIEHS Environmental Genome Project (EGP)
 - Candidate gene website
 - NIEHS web applications and other tools
 - GeneSNPs, PolyDoms, TraFac, PolyPhen, ECR Browser, GVS

International HapMap Project

Home | About the Project | Data | Publications | Conference

FR | English | Français | 日本語 | Yoruba

The International HapMap Project is a partnership of scientists and funding agencies from Canada, China, Japan, Nigeria, the United Kingdom and the United States to develop a public resource that will help researchers find genes associated with human disease and response to pharmaceuticals. See "About the International HapMap Project" for more information.

Project Information

- About the Project
- HapMap Publications
- HapMap Conference
- HapMap Meeting List
- HapMap Project Participants
- HapMap Mirror Site in Japan

Project Data

- Genomic Browser
- ENCODE Project
- Guidelines For Data Use

Useful Links

- HapMap Project Press Release
- NHGRI HapMap Page
- NCBI Variation Database (dbSNP)
- Japanese SNP Database (JSNP)

News

- 2005-03-01: HapMap public release #16. ATTN: The so-called Phase I data freeze which marks a major milestone of the project: a genotyped common SNP every 1kb in all populations under study. Data available for bulk download and graphical browsing. Summary of genotyped SNPs.
- 2005-02-28: HapMap News Volume 1, 2004. This is the first in a series of newsletters to be published by the Coriell Institute for Medical Research to inform communities how their samples are being used. Each issue of the newsletter will be available in the primary language of all the participating communities.
- 2005-02-07: International HapMap Consortium Expands Mapping Effort. The International HapMap Consortium, boosted by an additional 3.3 million in public-private support, announces plans to create an even more powerful map of human genetic variation than originally envisioned. The map will accelerate the discovery of genes related to common diseases, such as asthma, cancer, diabetes and heart disease.
- Old News

www.hapmap.org

Finding SNPs: HapMap Browser

[Show banner] [Hide instructions] [Bookmark this view] [Link to an image of this view] [Publication quality image] [Help]

Hap IDs: chr2:113682482..113689501 - Retrieving genotype data - Retrieving frequency data - Symbolic colours used:

Landmark or Region: Search Show 7.02 kb

Population descriptors: CEU: CEPH (Utah residents with ancestry from northern and western Europe), HCB: Han Chinese in Beijing, China, JPT: Japanese in Tokyo, Japan, YRI: Yoruba in Ibadan, Nigeria

rs1143643(-)
IL1B (JPT | YRI)

LocusLink genes
IL1B : Interleukin 1, beta precursor. The protein encoded by this gene...

RefSeq mRNAs
NM_002276

Finding SNPs: HapMap Browser

Home | About the Project | Data | Publications | Conference

SNP info:

rs1143643 with alleles A/G in dbSNP (dbSNP report | Ensembl SNPView)

Genotype frequency report

Population	Ref homozygote	Genotype frequencies heterozygote	Other homozygote	Allele frequencies													
				Total	Total												
	genotype	freq	count	allele	freq												
CEU	G/G	0.367	22	A/G	0.483	29	A/A	0.150	9	G	0.608	73	A	0.392	47	120	retrieve genotypes
HCB	G/G	0.222	10	A/G	0.533	24	A/A	0.244	11	G	0.489	44	A	0.511	45	90	retrieve genotypes
JPT	G/G	0.227	10	A/G	0.523	23	A/A	0.250	11	G	0.489	43	A	0.511	45	88	retrieve genotypes
YRI	G/G	0.700	42	A/G	0.300	18	A/A	0.000	0	G	0.800	102	A	0.150	18	120	retrieve genotypes

Note: the reference allele is the base observed in the reference genome sequence at this location.

Population descriptors:
CEU: CEPH (Utah residents with ancestry from northern and western Europe)
HCB: Han Chinese in Beijing, China
JPT: Japanese in Tokyo, Japan
YRI: Yoruba in Ibadan, Nigeria

Please see this page for more information about the populations, as well as a general discussion of the populations under study in the project.

Finding SNPs: HapMap Genotypes

```

#Sun Apr 10 16:01:28 2005: HapMap frequency data dump: 3 SNPs genotyped in population CEU on chr2:113683448..113693448
#see details on file format: see http://www.hapmap.org/genotype/
chr 2:113683448 chr2:113683448 chr2:113683448 chr2:113683448 chr2:113683448 chr2:113683448 chr2:113683448 chr2:113683448 chr2:113683448 chr2:113683448
rs1143643 rs1143643 rs1143643 rs1143643 rs1143643 rs1143643 rs1143643 rs1143643 rs1143643 rs1143643
AA GA GG AA GA GG AA GA GG AA GA GG AA GA GG AA GA GG AA GA GG AA GA GG AA GA GG
    
```

Bulk data downloads

- Genotypes: Individual genotype data submitted to the DCC to date.
- LD Data: Linkage disequilibrium properties D', LOD, R² compiled from the genotype data to date
- Allocated SNPs: dbSNP reference SNP clusters that have been picked and prioritized for genotyping according to several criteria (see info on how SNPs were selected). The file DOREADME contains per-chromosome SNP counts and further details.
- Frequencies: Allele & genotype frequencies compiled from genotyping data submitted to the DCC to date. These have also been submitted to dbSNP and should be available in the next dbSNP build.
- SNP assays: Details about assays submitted to the DCC to date. PCR primers, extension probes etc., specific to each genotyping platform.
- Protocols: Information on assay design, genotyping and other protocols used in the project.
- Samples/Individuals: Information on the samples used in the project and the individuals from which they were drawn. (See About the project: Which Populations are Being Sampled).
- XML docs: Documentation on the XML format used in the project.

Finding SNPs: HapMap Browser

For performing in-depth LD and haplotype analysis of genotype data install Haploview (ver.0.5) now available for download.

Data Source: HapMap Data Release 2 (Mar05), on NCBI B34 assembly, dbSNP b122

Highlight SNP Properties: Conigs Genotyped SNPs RefSeq mRNAs

CYP overview g/g SNPs/5000P Heterozygosity/1000P Sequence Tagged Sites

dbSNP SNPs/5000P Known genes/5000P SNP coverage/5000P

DNA/GC Content LocusLink genes SNP coverage/5000P

Gaps NT contigs?

Minimal SNP information for genotyping/characterization

- What is the SNP? Flanking sequence and alleles.
 - ✓ FASTA format

```

>snp_name
ACCGAGTACCCAG
[A/G]
ACTGGGATAGAAC
      
```

- dbSNP reference SNP # (rs #)
- Where is the SNP mapped? Exon, promoter, UTR, etc.
- ✓ picture of gene with mapped to the gene structure.
- How was it discovered? Method
- What assurances do you have that it is real? Validated how?
- What population – African, European, etc?
- What is the allele frequency of each SNP? Common (>10%), rare
- Are other SNPs associated - redundant? Genotyping data!

Finding SNPs: HapMap Browser

1. HapMap data sets are useful because individual genotype data can be used to determine optimal genotyping strategies (tagSNPs) or perform population genetic analyses (linkage disequilibrium)
2. Data are specific produced by those projects (not all dbSNP)
 - ✓ HapMap data is available in dbSNP
3. HapMap data (Phase II) can be accessed released prior to dbSNPs
4. Easier visualization of data and direct access to SNP data, individual genotypes, and LD analysis

Finding SNPs: Databases and Extraction

How do I find and download SNP data for analysis/genotyping?

1. Entrez Gene
 - dbSNP
 - Entrez SNP
2. HapMap Genome Browser
3. NIEHS Environmental Genome Project (EGP) Candidate gene website
4. NIEHS web applications and other tools
 - GeneSNPs, PolyDoms, TraFac, PolyPhen, ECR Browser, GVS

Finding SNPs: NIEHS SNPs Candidate Genes

Finding SNPs: NIEHS SNPs Candidate Genes

Finding SNPs: NIEHS SNPs Candidate Genes

Finding SNPs: NIEHS SNPs Candidate Genes

Download a zip file of all data for this gene	Sample Population Description
Mapping Data	cSNPs cDNA Color FASTA SNP Context PCR Primers (FASTA) GenBank
Genotyping Data	Visual Genotype Individual Genotypes SNP Alleles SNP Allele Frequency SNP Hardy-Weinberg
Haplotyping Data	PHASE Output Visual Haplotype Sorted by Frequency Sorted by Frequency
Linkage Data	LD Select (Tag SNPs) African Descent European Descent Hispanic Descent Asian Descent
Predictive Analyses	Nonsynonymous cSNP Analysis

Haplotyping Data PHASE Output Phased Individual Haplotypes Sorted by Frequency

Visual Haplotype

rsid	rsid	rsid	rsid
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
13	13	13	13
14	14	14	14
15	15	15	15
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21	21	21	21
22	22	22	22
23	23	23	23
24	24	24	24
25	25	25	25
26	26	26	26
27	27	27	27
28	28	28	28
29	29	29	29
30	30	30	30
31	31	31	31
32	32	32	32
33	33	33	33
34	34	34	34
35	35	35	35
36	36	36	36

LD Linkage Data LD Select (Tag SNPs)

African Descent European Descent Hispanic Descent Asian Descent

Bin	total_sites	average_misorder_allele_frequency
Bin 1	13	294
Bin 2	4	394
Bin 3	4	234
Bin 4	3	334
Bin 5	3	84
Bin 6	2	494
Bin 7	2	264

Predictive Analyses Nonsynonymous cSNP Analysis

rsid	AllelePop	a.a. Pos	Residue	Variant	Freq	PPHpredict	SIFTpredict
rsid1	03121	275	G	R	0.01	benign	TOLERATED
rsid1	031875	356	O	R	0.01	possibly damaging	INTOLERANT
rsid1	032885	693	D	N	0.05	benign	TOLERATED
rsid1	033426	871	P	L	0.43	benign	TOLERATED
rsid1	033921	1038	E	G	0.32	benign	INTOLERANT
rsid1	033927	1040	S	N	0.01	benign	TOLERATED
rsid1	034226	1140	S	G	0.02	benign	TOLERATED
rsid1	034356	1183	K	R	0.33	benign	TOLERATED
rsid1	055298	1613	S	G	0.33	benign	TOLERATED
rsid1	055319	1620	T	A	0.01	benign	TOLERATED

SIFT = Sorting Intolerant From Tolerant
Evolutionary comparison of non-synonymous SNPs

PolyPhen - Polymorphism Phenotyping
Structural protein characteristics and evolutionary comparison

Finding SNPs: NIEHS SNPs Candidate Genes

Download a zip file of all data for this gene

Mapping Data	cSNPs	Color FASTA	PCR Primers (FASTA)
Genotyping Data	Visual Genotype	SNP Alleles	SNP Hardy-Weinberg
Haplotyping Data	PHASE Output	Phased individual Haplotypes	Sorted by Frequency
Linkage Data	LD Select (Tag SNPs)	African Descent	European Descent
Predictive Analyses	Nonsynonymous cSNP Analysis	Hispanic Descent	Asian Descent

Finding SNPs: NIEHS SNPs Candidate Genes

National Institute of Environmental Health Sciences Environmental Genome Project

NIEHS SNPs

Welcome to the NIEHS SNPs Program

Introduction

The NIEHS Environmental Genome Project is a multi-disciplinary, collaborative effort focused on exploring the relationships between environmental exposures, gene-environment interactions, and human health outcomes. The NIEHS SNPs Program is a sub-project of the Environmental Genome Project that focuses on identifying common sequence variants (SNPs) in human genes involved in DNA repair and cell cycle pathways (see the article "Genetics in the repair and maintenance of the cell"). Identifying the SNPs that provide genetic maps of human genes that can be applied in evaluating human disease risk with environmental exposures.

Genes/SNPs Database

NIEHS SNPs are available in the Genes/SNPs database, as well as the national database resources, dbSNP, Gene/SNPs, and the National Center for Human Genome Research. The NIEHS SNPs Program also provides a gene-specific database of SNPs that are associated with the NIEHS. This database provides gene-specific information on SNPs that are associated with human disease and polymorphisms in select populations. This information is key in selecting the polymorphic sites needed to examine disease risk in human population studies.

Polymorphism Analysis

Automated DNA sequencing is being used to identify and genotype SNPs in human candidate genes (see PolyPhen). Candidate genes are being presented to identify common sequence variants for functional analysis and population-based studies. Candidate genes are normally sequenced across a population of individuals. The SNPs that are identified are then compared to the dbSNP database to determine if they are known SNPs.

egp.gs.washington.edu

Finding SNPs: NIEHS SNPs Candidate Genes

Data Downloads

All NIEHS SNPs Variation Data

Full Download of All Variation Data Files

WARNING: This is a large file and will take several minutes to download. The file is a compressed and "zipped" file containing the entire directory of text files. Please use the same text files which appear in the data pages for each candidate gene in our Project Gene Table. Please use our directory if this will be used in a publication.

Download of Variation Data (Single File)

Global Positioning File

This is a tab-delimited text file in our "bed" format that describes all SNPs discovered by NIEHS SNPs. The format of this file is:

Line format:
chromosome:position:chromosome:HUGO_NAME > cSNP Sample ID > Allele1 > Allele2

Example: 955484-REV1L2.D107.G.G

The chromosome position is generated from mapping to the most recent genome assembly available from the USCS Genome Browser.

Download of EGP Variation Data by Chromosome

These are tab-delimited text files in our "bed" format that describe all SNPs discovered by NIEHS SNPs but repeated info based on chromosome. The format of this file is:

NIEHS SNPs SIFT/PolyPhen Data

Functional changes in a candidate gene's protein function were assessed by taking the nonsynonymous coding SNPs (cSNPs) for each gene and using SIFT and PolyPhen. Generally, each nonsynonymous amino acid change is evaluated in the context of other available amino acid changes to determine the likelihood of the polymorphic nonsynonymous change and then statistically classified. These programs classify each coding SNP as benign or tolerant (SIFT) or as benign, probably damaging, or intolerant (PolyPhen).

Continued SIFT/PolyPhen Data for NIEHS SNPs Nonsynonymous SNPs (Benign or Probably Damaging)

Continued SIFT/PolyPhen Data for NIEHS SNPs Nonsynonymous SNPs (Intolerant or Probably Damaging)

SIFT Data for NIEHS SNPs Nonsynonymous SNPs (Probably Benign)

PolyPhen Data for NIEHS SNPs Nonsynonymous SNPs (Probably Damaging)

Finding SNPs: Databases and Extraction

How do I find and download SNP data for analysis/genotyping?

1. Entrez Gene
 - dbSNP
 - Entrez SNP
2. HapMap Genome Browser
3. NIEHS Environmental Genome Project (EGP)
Candidate gene website
4. NIEHS web applications and other tools
GeneSNPs, PolyDoms, TraFac, PolyPhen, ECR Browser, GVS

GeneSNPs

Graphic view of SNPs in context of gene elements

All NIEHS genes presented

- organized by pathway/function
- SNPs from dbSNP

- organized by submitter handle

Sequence context of SNPs presented in Color Fasta format

Link-outs to EntrezSNP pages

Summary "Genome SNPs" internal SNP viewer for one-stop SNP shopping

<http://www.genome.utah.edu/genesnps/>

GeneSNPs: One stop shopping

Gene	SNP ID	Position	Alleles	Frequency	Submitter
ADH1B	rs1044396	100,000,000	A/G	0.1	dbSNP
ADH1B	rs1044397	100,000,000	A/G	0.1	dbSNP
ADH1B	rs1044398	100,000,000	A/G	0.1	dbSNP
ADH1B	rs1044399	100,000,000	A/G	0.1	dbSNP
ADH1B	rs1044400	100,000,000	A/G	0.1	dbSNP
ADH1B	rs1044401	100,000,000	A/G	0.1	dbSNP
ADH1B	rs1044402	100,000,000	A/G	0.1	dbSNP
ADH1B	rs1044403	100,000,000	A/G	0.1	dbSNP
ADH1B	rs1044404	100,000,000	A/G	0.1	dbSNP
ADH1B	rs1044405	100,000,000	A/G	0.1	dbSNP

GeneSNPs: One stop shopping

Gene	SNP ID	Position	Alleles	Frequency	Submitter
ADH1B	rs1044396	100,000,000	A/G	0.1	dbSNP

GeneSNPs: One stop shopping

Gene	SNP ID	Position	Alleles	Frequency	Submitter
ADH1B	rs1044396	100,000,000	A/G	0.1	dbSNP

Polydoms

A web-based application that maps synonymous and non-synonymous SNPs onto known functional protein domains

- SNPs are from dbSNP and GeneSNPs
- Domain structures from NCBI's Conserved Domain Database
- Functional predictions based on SIFT and PolyPhen
- 3 dimensional mapping of SNPs on protein structure using Chime viewer

<http://polydoms.cchmc.org/polydoms/>

Polydoms

Polydoms

Mapping of nsSNPs onto protein structure

TraFac: Transcription Factor Binding Site Comparison

A tool for validating cis regulatory elements conserved between human and mouse

- Aligns human and mouse sequences using BLASTZ
- Consensus transcription factor binding sequences from Transfac database

<http://trafac.cchmc.org/trafac>

TraFac: Transcription Factor Binding Site Comparison

All TFBS in common

TFBS in parallel

ECR Browser: Evolutionary Conserved Regions

Aligns sequences to Mouse, Rat, Dog, Opposum, Chicken, Fugu and Drosophila

Gene annotations from UCSC Genome Browser

Easy retrieval of ECR sequences and alignments

Pre-computed transcription factor binding sites

<http://ecrbrowser.dcode.org>

ECR Browser: Evolutionary Conserved Regions

GVS: Genome Variation Server beta Sponsored by SeattleSNPs

Gene Name: NOS2A
Gene ID: 4982
Chromosome 17: 23187919 - 23151602 (-)

Select Population(s)
Check at least one population. You may select up to 4 populations. All those selected have individuals in common.

Select	Number Variants	Number Individuals	Population	Subgroup	Check to Select Only Variants in Common
<input type="checkbox"/>	235	2418	EGP_Yorub-PANEL	EGP_BNFS	
<input type="checkbox"/>	235	3816	EGP_ASIAN-PANEL	EGP_BNFS	
<input type="checkbox"/>	235	3168	EGP_KO-PANEL	EGP_BNFS	
<input type="checkbox"/>	235	5153	EGP_CEPH-PANEL	EGP_BNFS	
<input type="checkbox"/>	269	4176	EGP_HCB-PANEL	EGP_BNFS	
<input type="checkbox"/>	51	4182	HapMap-CEP	CEPH-HAPMAP	<input type="checkbox"/>
<input type="checkbox"/>	50	2250	HapMap-JPT	CEPH-HAPMAP	<input type="checkbox"/>
<input type="checkbox"/>	50	2250	HapMap-MOZ	CEPH-HAPMAP	<input type="checkbox"/>
<input type="checkbox"/>	48	4320	HapMap-CEU	CEPH-HAPMAP	<input type="checkbox"/>
<input type="checkbox"/>	21	504	AFR_CEPH-PANEL	CEPH-CEP	<input type="checkbox"/>
<input type="checkbox"/>	21	483	AFR_JPT-PANEL	CEPH-CEP	<input type="checkbox"/>
<input type="checkbox"/>	21	504	AFR_KO-PANEL	CEPH-CEP	<input type="checkbox"/>

Set Parameters

Output SNPs By	SNP_Position	Allele Frequency Cutoff (%)	5
Display SNPs By	Tab/Matrix	r ² Threshold (0.5-1.0)	0.8
Cluster SNPs	<input type="checkbox"/>	LD Minimum (0.5-1.0)	0.7
Cluster Samples	<input type="checkbox"/>	LD Maximum (0.5-1.0)	0.9

Display Results

Navigation: [Home](#) [Help](#) [About](#) [Feedback](#) [Privacy Policy](#) [Terms of Use](#)

GVS: Genome Variation Server beta Sponsored by SeattleSNPs

Display Results

Navigation: [Home](#) [Help](#) [About](#) [Feedback](#) [Privacy Policy](#) [Terms of Use](#)

EGP Yoruban population

VQ2 of: Gene Name NOS2A

UTR and Coding SNPs

Gene Name: NOS2A
Allele Frequency Cutoff (%) 5
r² Threshold 0.8
Population: EGP_Yorub-PANEL, Subgroup: EGP_BNFS

rs ID	SNP Position	Allele Frequency (%)	Tag SNPs	Other SNPs	
1	10	12.4	2312079	2311270, 2311271, 2311272, 2311273, 2311274, 2311275, 2311276, 2311277, 2311278, 2311279, 2311280, 2311281, 2311282, 2311283, 2311284, 2311285, 2311286, 2311287, 2311288, 2311289, 2311290, 2311291, 2311292, 2311293, 2311294, 2311295, 2311296, 2311297, 2311298, 2311299, 2311300, 2311301, 2311302, 2311303, 2311304, 2311305, 2311306, 2311307, 2311308, 2311309, 2311310, 2311311, 2311312, 2311313, 2311314, 2311315, 2311316, 2311317, 2311318, 2311319, 2311320, 2311321, 2311322, 2311323, 2311324, 2311325, 2311326, 2311327, 2311328, 2311329, 2311330, 2311331, 2311332, 2311333, 2311334, 2311335, 2311336, 2311337, 2311338, 2311339, 2311340, 2311341, 2311342, 2311343, 2311344, 2311345, 2311346, 2311347, 2311348, 2311349, 2311350, 2311351, 2311352, 2311353, 2311354, 2311355, 2311356, 2311357, 2311358, 2311359, 2311360, 2311361, 2311362, 2311363, 2311364, 2311365, 2311366, 2311367, 2311368, 2311369, 2311370, 2311371, 2311372, 2311373, 2311374, 2311375, 2311376, 2311377, 2311378, 2311379, 2311380, 2311381, 2311382, 2311383, 2311384, 2311385, 2311386, 2311387, 2311388, 2311389, 2311390, 2311391, 2311392, 2311393, 2311394, 2311395, 2311396, 2311397, 2311398, 2311399, 2311400, 2311401, 2311402, 2311403, 2311404, 2311405, 2311406, 2311407, 2311408, 2311409, 2311410, 2311411, 2311412, 2311413, 2311414, 2311415, 2311416, 2311417, 2311418, 2311419, 2311420, 2311421, 2311422, 2311423, 2311424, 2311425, 2311426, 2311427, 2311428, 2311429, 2311430, 2311431, 2311432, 2311433, 2311434, 2311435, 2311436, 2311437, 2311438, 2311439, 2311440, 2311441, 2311442, 2311443, 2311444, 2311445, 2311446, 2311447, 2311448, 2311449, 2311450, 2311451, 2311452, 2311453, 2311454, 2311455, 2311456, 2311457, 2311458, 2311459, 2311460, 2311461, 2311462, 2311463, 2311464, 2311465, 2311466, 2311467, 2311468, 2311469, 2311470, 2311471, 2311472, 2311473, 2311474, 2311475, 2311476, 2311477, 2311478, 2311479, 2311480, 2311481, 2311482, 2311483, 2311484, 2311485, 2311486, 2311487, 2311488, 2311489, 2311490, 2311491, 2311492, 2311493, 2311494, 2311495, 2311496, 2311497, 2311498, 2311499, 2311500, 2311501, 2311502, 2311503, 2311504, 2311505, 2311506, 2311507, 2311508, 2311509, 2311510, 2311511, 2311512, 2311513, 2311514, 2311515, 2311516, 2311517, 2311518, 2311519, 2311520, 2311521, 2311522, 2311523, 2311524, 2311525, 2311526, 2311527, 2311528, 2311529, 2311530, 2311531, 2311532, 2311533, 2311534, 2311535, 2311536, 2311537, 2311538, 2311539, 2311540, 2311541, 2311542, 2311543, 2311544, 2311545, 2311546, 2311547, 2311548, 2311549, 2311550, 2311551, 2311552, 2311553, 2311554, 2311555, 2311556, 2311557, 2311558, 2311559, 2311560, 2311561, 2311562, 2311563, 2311564, 2311565, 2311566, 2311567, 2311568, 2311569, 2311570, 2311571, 2311572, 2311573, 2311574, 2311575, 2311576, 2311577, 2311578, 2311579, 2311580, 2311581, 2311582, 2311583, 2311584, 2311585, 2311586, 2311587, 2311588, 2311589, 2311590, 2311591, 2311592, 2311593, 2311594, 2311595, 2311596, 2311597, 2311598, 2311599, 2311600, 2311601, 2311602, 2311603, 2311604, 2311605, 2311606, 2311607, 2311608, 2311609, 2311610, 2311611, 2311612, 2311613, 2311614, 2311615, 2311616, 2311617, 2311618, 2311619, 2311620, 2311621, 2311622, 2311623, 2311624, 2311625, 2311626, 2311627, 2311628, 2311629, 2311630, 2311631, 2311632, 2311633, 2311634, 2311635, 2311636, 2311637, 2311638, 2311639, 2311640, 2311641, 2311642, 2311643, 2311644, 2311645, 2311646, 2311647, 2311648, 2311649, 2311650, 2311651, 2311652, 2311653, 2311654, 2311655, 2311656, 2311657, 2311658, 2311659, 2311660, 2311661, 2311662, 2311663, 2311664, 2311665, 2311666, 2311667, 2311668, 2311669, 2311670, 2311671, 2311672, 2311673, 2311674, 2311675, 2311676, 2311677, 2311678, 2311679, 2311680, 2311681, 2311682, 2311683, 2311684, 2311685, 2311686, 2311687, 2311688, 2311689, 2311690, 2311691, 2311692, 2311693, 2311694, 2311695, 2311696, 2311697, 2311698, 2311699, 2311700, 2311701, 2311702, 2311703, 2311704, 2311705, 2311706, 2311707, 2311708, 2311709, 2311710, 2311711, 2311712, 2311713, 2311714, 2311715, 2311716, 2311717, 2311718, 2311719, 2311720, 2311721, 2311722, 2311723, 2311724, 2311725, 2311726, 2311727, 2311728, 2311729, 2311730, 2311731, 2311732, 2311733, 2311734, 2311735, 2311736, 2311737, 2311738, 2311739, 2311740, 2311741, 2311742, 2311743, 2311744, 2311745, 2311746, 2311747, 2311748, 2311749, 2311750, 2311751, 2311752, 2311753, 2311754, 2311755, 2311756, 2311757, 2311758, 2311759, 2311760, 2311761, 2311762, 2311763, 2311764, 2311765, 2311766, 2311767, 2311768, 2311769, 2311770, 2311771, 2311772, 2311773, 2311774, 2311775, 2311776, 2311777, 2311778, 2311779, 2311780, 2311781, 2311782, 2311783, 2311784, 2311785, 2311786, 2311787, 2311788, 2311789, 2311790, 2311791, 2311792, 2311793, 2311794, 2311795, 2311796, 2311797, 2311798, 2311799, 2311800, 2311801, 2311802, 2311803, 2311804, 2311805, 2311806, 2311807, 2311808, 2311809, 2311810, 2311811, 2311812, 2311813, 2311814, 2311815, 2311816, 2311817, 2311818, 2311819, 2311820, 2311821, 2311822, 2311823, 2311824, 2311825, 2311826, 2311827, 2311828, 2311829, 2311830, 2311831, 2311832, 2311833, 2311834, 2311835, 2311836, 2311837, 2311838, 2311839, 2311840, 2311841, 2311842, 2311843, 2311844, 2311845, 2311846, 2311847, 2311848, 2311849, 2311850, 2311851, 2311852, 2311853, 2311854, 2311855, 2311856, 2311857, 2311858, 2311859, 2311860, 2311861, 2311862, 2311863, 2311864, 2311865, 2311866, 2311867, 2311868, 2311869, 2311870, 2311871, 2311872, 2311873, 2311874, 2311875, 2311876, 2311877, 2311878, 2311879, 2311880, 2311881, 2311882, 2311883, 2311884, 2311885, 2311886, 2311887, 2311888, 2311889, 2311890, 2311891, 2311892, 2311893, 2311894, 2311895, 2311896, 2311897, 2311898, 2311899, 2311900, 2311901, 2311902, 2311903, 2311904, 2311905, 2311906, 2311907, 2311908, 2311909, 2311910, 2311911, 2311912, 2311913, 2311914, 2311915, 2311916, 2311917, 2311918, 2311919, 2311920, 2311921, 2311922, 2311923, 2311924, 2311925, 2311926, 2311927, 2311928, 2311929, 2311930, 2311931, 2311932, 2311933, 2311934, 2311935, 2311936, 2311937, 2311938, 2311939, 2311940, 2311941, 2311942, 2311943, 2311944, 2311945, 2311946, 2311947, 2311948, 2311949, 2311950, 2311951, 2311952, 2311953, 2311954, 2311955, 2311956, 2311957, 2311958, 2311959, 2311960, 2311961, 2311962, 2311963, 2311964, 2311965, 2311966, 2311967, 2311968, 2311969, 2311970, 2311971, 2311972, 2311973, 2311974, 2311975, 2311976, 2311977, 2311978, 2311979, 2311980, 2311981, 2311982, 2311983, 2311984, 2311985, 2311986, 2311987, 2311988, 2311989, 2311990, 2311991, 2311992, 2311993, 2311994, 2311995, 2311996, 2311997, 2311998, 2311999, 2312000	

Variation Color Code: Coding (red), UTR (orange)

GVS: Genome Variation Server beta Sponsored by SeattleSNPs

Display Results

Navigation: [Home](#) [Help](#) [About](#) [Feedback](#) [Privacy Policy](#) [Terms of Use](#)

HapMap CEPH population

VQ2 of: Gene Name NOS2A

Gene Name: NOS2A
Allele Frequency Cutoff (%) 5
r² Threshold 0.8
Population: HapMap-CEP, Subgroup: HAPMAP

SNP	Position	Allele	Frequency (%)
2310888	1420107	A	0.000000
2310889	1420108	C	0.000000
2310890	1420109	A	0.000000
2310891	1420110	C	0.000000
2310892	1420111	A	0.000000
2310893	1420112	C	0.000000
2310894	1420113	A	0.000000
2310895	1420114	C	0.000000
2310896	1420115	A	0.000000
2310897	1420116	C	0.000000
2310898	1420117	A	0.000000
2310899	1420118	C	0.000000
2310900	1420119	A	0.000000
2310901	1420120	C	0.000000
2310902	1420121	A	0.000000
2310903	1420122	C	0.000000
2310904	1420123	A	0.000000
2310905	1420124	C	0.000000
2310906	1420125	A	0.000000
2310907	1420126	C	0.000000
2310908	1420127	A	0.000000
2310909	1420128	C	0.000000
2310910	1420129	A	0.000000
2310911	1420130	C	0.000000
2310912	1420131	A	0.000000
2310913	1420132	C	0.000000
2310914	1420133	A	0.000000
2310915	1420134	C	0.000000
2310916	1420135	A	0.000000
2310917	1420136	C	0.000000
2310918	1420137	A	0.000000
2310919	1420138	C	0.000000
2310920	1420139	A	0.000000
2310921	1420140	C	0.000000
2310922	1420141	A	0.000000
2310923	1420142	C	0.000000
2310924	1420143	A	0.000000
2310925	1420144	C	0.000000
2310926	1420145	A	0.000000
2310927	1420146	C	0.000000
2310928	1420147	A	0.000000
2310929	1420148	C	0.000000
2310930	1420149	A	0.000000
2310931	1420150	C	0.000000
2310932	1420151	A	0.000000
2310933	1420152	C	0.000000
2310934	1420153	A	0.000000
2310935	1420154	C	0.000000

Variation Color Code: Coding (red), UTR (orange)

GVS: Genome Variation Server beta Sponsored by SeattleSNPs

Display Results

Navigation: [Home](#) [Help](#) [About](#) [Feedback](#) [Privacy Policy](#) [Terms of Use](#)

VQ2 of: Gene Name NOS2A

Gene Name: NOS2A
Allele Frequency Cutoff (%) 5
r² Threshold 0.8
Population: EGP_Yorub-PANEL, Subgroup: EGP_BNFS

rs ID	SNP Position	Allele Frequency (%)	Tag SNPs	Other SNPs
1	10	12.4	2312079	2311270, 2311271, 2311272, 2311273, 2311274, 2311275, 2311276, 2311277, 2311278, 2311279, 2311280, 2311281, 2311282, 2311283, 2311284, 2311285, 2311286, 2311287, 2311288, 2311289, 2311290, 2311291, 2311292, 2311293, 2311294, 2311295, 2311296, 2311297, 2311298, 2311299, 2311300, 2311301, 2311302, 2311303, 2311304, 2311305, 2311306, 2311307, 2311308, 2311309, 2311310, 2311311, 2311312, 2311313, 2311314, 2311315, 2311316, 2311317, 2311318, 2311319, 2311320, 2311321, 2311322, 2311323, 2311324, 2311325, 2311326, 2311327, 2311328, 2311329, 2311330, 2311331, 2311332, 2311333, 2311334, 2311335, 2311336, 2311337, 2311338, 2311339, 2311340, 2311341, 2311342, 2311343, 2311344, 2311345, 2311346, 2311347, 2311348, 2311349, 2311350, 2311351, 2311352, 2311353, 2311354, 2311355, 2311356, 2311357, 2311358, 2311359, 2311360, 2311361, 2311362, 2311363, 2311364, 2311365, 2311366, 2311367, 2311368, 2311369, 2311370, 2311371, 2311372, 2311373, 2311374, 2311375, 2311376, 2311377, 2311378, 2311379, 2311380, 2311381, 2311382, 2311383, 2311384, 2311385, 2311386, 2311387, 2311388, 2311389, 2311390, 2311391, 2311392, 2311393, 2311394, 2311395, 2311396, 2311397, 2311398, 2311399, 2311400, 2311401, 2311402, 2311403, 2311404, 2311405, 2311406, 2311407, 2311408, 2311409, 2311410, 2311411, 2311412, 2311413, 2311414, 2311415, 2311416, 2311417, 2311418, 2311419, 2311420, 2311421, 2311422, 2311423, 2311424, 2311425, 2311426, 2311427, 23114

GVS: Genome Variation Server beta Sponsored by SeattleSNPs

Display Results

Gene Name: **NDR2A**
 Allele Frequency cutoff (r%) : 5
 Population: **POP_10000-PAN01, Submitter: EGP_RAP0**

SNP ID	Ref Allele	Alt Allele	Frequency (%)	rsID	High-Throughput Genotyping	Gene	Function
2118499	CC	TT	0.44	1355	NO	NDR2A	missense
21184113	CC	TT	0.41	690	NO	NDR2A	missense
21184635	AC	TC	0.23	823	NO	NDR2A	missense
21184687	AC	TC	0.23	823	NO	NDR2A	missense
21184756	AC	TC	0.23	823	NO	NDR2A	missense
21184815	AC	TC	0.23	823	NO	NDR2A	missense
21184840	AC	TC	0.23	823	NO	NDR2A	missense
21184841	AC	TC	0.23	823	NO	NDR2A	missense
21184842	AC	TC	0.23	823	NO	NDR2A	missense
21184843	AC	TC	0.23	823	NO	NDR2A	missense
21184844	AC	TC	0.23	823	NO	NDR2A	missense
21184845	AC	TC	0.23	823	NO	NDR2A	missense
21184846	AC	TC	0.23	823	NO	NDR2A	missense
21184847	AC	TC	0.23	823	NO	NDR2A	missense
21184848	AC	TC	0.23	823	NO	NDR2A	missense
21184849	AC	TC	0.23	823	NO	NDR2A	missense
21184850	AC	TC	0.23	823	NO	NDR2A	missense
21184851	AC	TC	0.23	823	NO	NDR2A	missense
21184852	AC	TC	0.23	823	NO	NDR2A	missense
21184853	AC	TC	0.23	823	NO	NDR2A	missense
21184854	AC	TC	0.23	823	NO	NDR2A	missense
21184855	AC	TC	0.23	823	NO	NDR2A	missense
21184856	AC	TC	0.23	823	NO	NDR2A	missense
21184857	AC	TC	0.23	823	NO	NDR2A	missense
21184858	AC	TC	0.23	823	NO	NDR2A	missense
21184859	AC	TC	0.23	823	NO	NDR2A	missense
21184860	AC	TC	0.23	823	NO	NDR2A	missense
21184861	AC	TC	0.23	823	NO	NDR2A	missense
21184862	AC	TC	0.23	823	NO	NDR2A	missense
21184863	AC	TC	0.23	823	NO	NDR2A	missense
21184864	AC	TC	0.23	823	NO	NDR2A	missense
21184865	AC	TC	0.23	823	NO	NDR2A	missense
21184866	AC	TC	0.23	823	NO	NDR2A	missense
21184867	AC	TC	0.23	823	NO	NDR2A	missense
21184868	AC	TC	0.23	823	NO	NDR2A	missense
21184869	AC	TC	0.23	823	NO	NDR2A	missense
21184870	AC	TC	0.23	823	NO	NDR2A	missense
21184871	AC	TC	0.23	823	NO	NDR2A	missense
21184872	AC	TC	0.23	823	NO	NDR2A	missense
21184873	AC	TC	0.23	823	NO	NDR2A	missense
21184874	AC	TC	0.23	823	NO	NDR2A	missense
21184875	AC	TC	0.23	823	NO	NDR2A	missense
21184876	AC	TC	0.23	823	NO	NDR2A	missense
21184877	AC	TC	0.23	823	NO	NDR2A	missense
21184878	AC	TC	0.23	823	NO	NDR2A	missense
21184879	AC	TC	0.23	823	NO	NDR2A	missense
21184880	AC	TC	0.23	823	NO	NDR2A	missense
21184881	AC	TC	0.23	823	NO	NDR2A	missense
21184882	AC	TC	0.23	823	NO	NDR2A	missense
21184883	AC	TC	0.23	823	NO	NDR2A	missense
21184884	AC	TC	0.23	823	NO	NDR2A	missense
21184885	AC	TC	0.23	823	NO	NDR2A	missense
21184886	AC	TC	0.23	823	NO	NDR2A	missense
21184887	AC	TC	0.23	823	NO	NDR2A	missense
21184888	AC	TC	0.23	823	NO	NDR2A	missense
21184889	AC	TC	0.23	823	NO	NDR2A	missense
21184890	AC	TC	0.23	823	NO	NDR2A	missense
21184891	AC	TC	0.23	823	NO	NDR2A	missense
21184892	AC	TC	0.23	823	NO	NDR2A	missense
21184893	AC	TC	0.23	823	NO	NDR2A	missense
21184894	AC	TC	0.23	823	NO	NDR2A	missense
21184895	AC	TC	0.23	823	NO	NDR2A	missense
21184896	AC	TC	0.23	823	NO	NDR2A	missense
21184897	AC	TC	0.23	823	NO	NDR2A	missense
21184898	AC	TC	0.23	823	NO	NDR2A	missense
21184899	AC	TC	0.23	823	NO	NDR2A	missense
21184900	AC	TC	0.23	823	NO	NDR2A	missense

Finding SNPs: Databases and Extraction

- One stop shopping
 - NIEHS SNPs and GeneSNPs
- Prediction of functional variations
 - Polydoms and PolyPhen
- Identification of transcription factor binding sites in Evolutionary Conserved Regions
 - TraFac and the ECR browser
- Visualization and analysis of LD and TagSNPs
 - GVS