# 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**

"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 ACCGAGTAGCCAG [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?

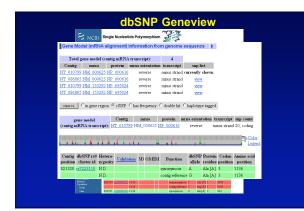
- 1. Entrez Gene
  - dbSNP
  - Entrez SNP
- 2. HapMap Genome Browser
- 3. NIEHS Environmental Genome Project (EGP) Candidate gene website
- 4. NIEHS web applications and other tools GeneSNPS, PolyDoms, TraFac, PolyPhen,

| Publied All Databas  | atabases BLAST OMIM Books Tax   | Information<br>nues of Health<br>Browser Structure   |
|--|---|--|
| SITE MAP<br>Alphabetical List<br>Resource Guide<br>About NCBI<br>An introduction to<br>NCBI<br>GenBank | What does NCBI do?     Established in 1988 as a national resource for<br>molecular biology information, NCBI creates<br>public databases, conduct sreasers in<br>computational biology, develops software<br>tools for analyzing genome data, and<br>dissernitates biomedical information - all for | Hot Spots  Assembly Archive  Custers of orthologous groups  Coffee Break, Genns & Disease. |
| Sequence<br>submission support<br>and software<br>Literature<br>databases<br>PubMed, OMIM,             | the better understanding of molecular<br>processes affecting human health and<br>disease. <u>More</u>   | NCBI Handbook<br>Electronic PCR<br>Entrez Home   |
| Books, and<br>PubMed Central<br>Molecular<br>databases<br>Sequences,                                   | comparison of influenza influenza influenza and<br>provides a reference for wiral sequences.<br>The resource contains data from the NAID<br>Influenza Genome Sequencing Project and<br>GenBank, Sa well as pre-competed<br>alignments of flu sequences  | Entrez Tools     Gene expression<br>omnibus (GEO)     Human genome<br>resources            |
| structures, and taxonomy   | Entrez Gene   | Malaria genetics &   |
| Genomic<br>biology   | You can now use Entrez to search for<br>information centered on the concept of a<br>gene, and connect to many sources of related<br>information both within and purside NCBL  | genomics<br>Map Viewer   |

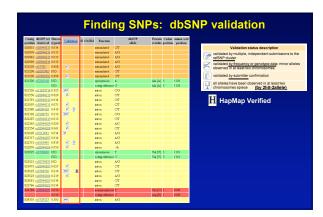


| NCBI - E |  | ng SNPs: Where do I start  | ?                                |
|----------|--|--|----------------------------------|
|          | S NCBI   | Entrez Gene  | My MCDI N<br>Right of Respirated |
|          | Search Gene  | Added Policities Policity Occurs Studies Plac Televiery  | current records only             |
|          | Entrez Gene<br>Home<br>About   | Linits Previewindes History Claptoand Datais<br>Deglay Summay Elberr (2) E Send to E<br>All: 14 Cennes Genomes: 13 S2P Cennetiew: 10 (2)   |                                  |
|          | FAQ<br>Help<br>Search Toolbar<br>Gene Handbook<br>Statistics<br>Downloads<br>(FTP) | Brune 1 - 1 of 14<br>C DBiological Synche INOSZA and Nume state order synthese 2A (solveb)e, hep-streyten) [Jörew agrice<br>Other Alasses: IEE MOOR, BIOS, BIOS2<br>Other Designations: 1003, pp. 02, non-onder synthese 2A, note order synthese, narrophage<br>Characterizer 1, Locientin, 17(11) 2, solved and 2, note order synthese materials. | One page.<br>Links               |
|          | Mailing Lists<br>Gene<br>RefSeq<br>Feedback  | Other Manuarian (a)     Other Manuarian (A. (aduchle, hepatocytee) [Caster,forstharde]     Other Alanses: INOS     Other Devanises: induchle mitic code conducts (dir): code contacts (A.  | Lines                            |
|          | Help Desk<br>Corrections<br>About GeneRIFs<br>Belated Stas                         | Chromesome: 9<br>GenerID: 408822   | Links                            |
|          | BLAST<br>Britiez Genome<br>Genome Projects<br>Genomic Biology<br>GDO               | mite existe sprahae 2.4 (adachth, hepatospra) [Gailiar gailine]<br>Oller Allace: DNOS<br>Oller Designations: Italachth mite code synfaure<br>Chronoscore: 19<br>GranD: 355807  |                                  |

| Finding SNPs: Entrez (  | Jene                      |              |
|---|---------------------------|--------------|
| S NCBI Entrez Gene  | Table Of Contents         | az<br>pater) |
| Search Gene Sort  | Conserved Domains         | 1.4          |
| Lints Preventer History Cichoard Defails  | GEO Profiles              | ł –          |
|   | HomoloGene<br>Map Viewer  |              |
| Augury Full Report 🔹 Stever 20 💌 Send to 💌  | Map Viewer<br>Nucleotide  |              |
| All: 1 Genes Genomes: 1 SNP GeneView: 1 🕱   | OMIM                      |              |
| L NOS2A nitric oxide synthase 2A (inducible, hepatocytes) [Homo supress]  | Full text in PMC<br>Prohe |              |
| SeneID: 4843 Primary source HGNC 7873   | Probe<br>Protein          | 2006         |
| Summary   | PubMed                    |              |
| fficial Symbol: NOS2A and Name: nitic oxide synthese 2A (inducible, hepatocytes) provided by HUGO G   | RubMed (GeneRIE)          |              |
| for related HPRD 01225 MIM 163730   | SNP<br>SNP: Genotype      |              |
| Gene type: protein coding   | SNP: GeneView             |              |
| iene name: NOS2A  | Tatonomy                  |              |
| Sene description: nitric oxide synthase 2A (inducible, hepatocytes)   | UnISTS                    |              |
| RefSeq status: Reviewed<br>Juganism: Howe capters   | CCDS                      |              |
| n genastin <u>150000 captente</u><br>ineesse: Bakarvota: Metazoa: Chordata: Craniata: Vertebrata: Bateleostomi: Mammalia: Batheria: Bua                           | Ensembl                   |              |
| Honoge, Index your, Mentalou, Colorada, Colorada, Versionad, Indexedición, Indexedición, Indexedición, Indexed<br>Hominidae: Homo                                 | Evidence Viewer           |              |
| Sene alianes: NOS, INOS, NOS2, HEP-NOS  | GDB<br>HONC               |              |
| Summary: Nitric oxide is a reactive free radical which acts as a biologic mediator in several processes, including (  | HPRD                      |              |
| ntitumoral activities. This gene encodes a nitric oxide synthase which is expressed in liver and is inducible by a cor  | KEGG                      | tain         |
| ytokines. Three related pseudogenes are located within the Smith-Magenia syndrome region on chromosome 17.<br>wo transcript variants encoding different isoforms. | ModelMaker<br>Pharmôk/9   | m            |
|   | PharmOKB<br>UCSC          |              |
| Genomic regions, transcripts, and products  | UniGene                   |              |
| ininus strand) RetSea below   | LinkOut                   |              |
| NC_000017   | 1 Entrez Gene Info        |              |
| P F   | H Feedback                | i l          |
| M. 19122  | * Subscriptions           |              |



|             |                             |                              | dbSN                                     | P G                | enevi           | ew    |                   |  |
|-------------|-----------------------------|------------------------------|--|--------------------|-----------------|-------|-------------------|--|
|             |                             |                              | leotide Polymorph                        | 10 Sec. 1          | ž               |       |                   |  |
| Gene Mo     | del (mRNA                   | alignment)                   | information fro                          | m genome           | sequence        |       |                   |  |
| Total e     | ene model (c                | ontig mBN4                   | transcript):                             | 4                  |                 |       |                   |  |
| Contig      | mina                        |                              | mma orientation                          |                    | snp list        |       |                   |  |
|             | NM 000625                   |                              |  |                    | currently show  | n     |                   |  |
| NT 086865   | NM 000625                   | NP 000616                    | reverse                                  | minus strand       | view            |       |                   |  |
| NT 010799   | NM 153292                   | NP 695024                    | reverse                                  | minus strand       | view            |       |                   |  |
| NT 086865   | NM 153292                   | NP 695024                    | reverse                                  | minus strand       | view            |       |                   |  |
| gene        | model                       | Contig                       | has frequency f<br>mma<br>99 NM_000625 h | protein m          | rna orientation |       |                   |  |
|             | <b></b> .                   | <b>.</b>                     | . <b>.</b>                               |                    |                 |       |                   |  |
| position ch | SNP rs# Het<br>uster id zyg | osity Valida                 | tion 3D OMIM                             | Function           |                 | dlele | Codon<br>position |  |
|             | <u>8944217</u> 0.01         |                              |  | untranslate d      | C/T             |       |                   |  |
|             | <u>8944216</u> 0.01         | 15 BOTTON BROOM              |  | untranslated       | A/G             |       |                   |  |
|             | Netod<br>Netod<br>Detal     | 821029 m1578<br>821200 m3201 | 2 0.423 🕺                                | Celtras<br>Celtras | shoed ANG       |       |                   |  |
|             |                             |                              |  |                    |                 |       |                   |  |



| Contig | dbSNP 158  | Hetero- |            |             | OMEM    | Function         | diSNP  | Protein | Codon    | Amino acid |
|--------|------------|---------|------------|-------------|---------|------------------|--------|---------|----------|------------|
|        | cluster id |         | Vablas     | <u>m</u> 33 | ) OBIIM | Punction         | allele | residue | position | position   |
| 820851 | cs28944217 | 0.014   |            |             |         | untranslated     | C/T    |         |          |            |
| 820939 | rs28944216 | 0.015   |            |             |         | untranslated     | A/G    |         |          |            |
|        | 028944215  |         |            |             |         |                  | A/G    |         |          |            |
| 820990 | 028944214  | 0.015   |            |             |         | untranslated     | A/G    |         |          |            |
| 821029 | rs15782    | 0.428   | 8          |             |         |                  | A/G    |         |          |            |
|        | ci3201742  |         |            |             |         |                  | CT     |         |          |            |
| 821200 | rs16966522 | 0.499   | R          |             |         | untranslated     | C/T    |         |          |            |
| 821320 |            | N.D.    |            |             |         | synonymous       | A      | Ala [A] | 3        | 1136       |
|        |            | N.D.    |            |             |         | contig reference |        | Ala [A] | 3        | 1136       |
| 821526 | cs11653716 | 0.325   | 8K         |             |         | intron.          | C/G    |         |          |            |
|        | rs28944213 |         | 8          |             |         |                  | OT     |         |          |            |
| 821791 | 028944212  | 0.014   |            |             |         | intron           | CT     |         |          |            |
| 821907 | cs28944211 | 0.071   | 8          |             |         | intron.          | C/T    |         |          |            |
|        | n8090418   |         | Z U        |             |         |                  | OT     |         |          |            |
|        | 07208775   |         | 1.         |             |         |                  | C/G    |         | -        |            |
|        | c:28944210 |         |            |             |         |                  | C/T    |         |          |            |
|        | rs28944209 |         | 8          |             |         |                  | A/C    |         |          |            |
|        | 128944208  |         | X          |             |         |                  | CT     |         |          |            |
|        | rs7213691  |         | 8          |             |         |                  | A/0    |         |          |            |
|        | rs28944207 |         |            |             |         |                  | A/G    |         |          |            |
|        | ci6505445  |         | <u>×</u> 1 |             |         |                  | A/G    |         |          |            |
|        | cs28944206 |         | 8          |             |         |                  | -/A    |         |          |            |
| 822925 | rx3729662  |         |            |             |         |                  | т      | Val[V]  |          | 1110       |
|        |            | N.D.    |            |             |         | costig reference |        | Val[V]  | 3        | 1110       |
|        | cs28730631 |         |            |             |         |                  | A/G    |         |          |            |
|        | rs28944205 |         | 8          |             |         |                  | or     |         |          |            |
|        | 03794756   |         |            | H           |         |                  | C/T    |         |          |            |
|        | cs28944204 |         | 8          |             |         |                  | A/G    |         |          |            |
|        | rs28944203 |         |            |             |         |                  | OT     |         |          |            |
|        | 028944202  |         |            |             |         |                  | C/T    |         |          |            |
| 824184 | cs28344201 |         |            |             |         | nonsynonymous    |        | Cps [C] |          | 1009       |
|        |            | 0.016   |            |             |         | costig reference |        | Arg [R] | 1        | 1009       |
| 824310 | 03729727   | 0.305   | 81         |             |         | intron           | A/G    |         |          |            |

|                           | F  | ind            | ling                     | I SN                     | IPs:                                       | dbS               | NP                               | datal     | oase            |           |  |
|---------------------------|--|----------------|--------------------------|--------------------------|--|-------------------|----------------------------------|-----------|-----------------|-----------|--|
|                           | Fasta sequence (   | Legend         | )                        |                          |  |                   |                                  | t         |                 |           |  |
| Sut                       | >gnl dbSNP ss38342908  | 3jalleiePos    | =256 len=                | =511 tazid=              | 9606 alleles='C/\                          | G∥mo⊫Gen          | omic                             |           |                 |           |  |
| The sul                   | GGGAGGCACC CCCTTT  |                |                          |                          |  |                   |                                  |           |                 |           |  |
| NC<br>Assa                | GAGCCTAGAG TTCAGTC<br>TCTCCCTAAG ACACATT<br>GGGATAGAGC AGTAAAT<br>CCAGACAAGA AAGAG | CAT ACG        | TTCAACT                  | GAACATTT                 | AT TGAGCACCTA                              | CTATGTO           | TA                               |           |                 |           |  |
| <u>55108</u><br>51383-    | TAAATAGGGT GGCCAGG   | GAA GGC        | AGAGCTG                  | GAAGGTOG                 |  | AACCTGAG          | AT                               |           |                 |           |  |
| Fas                       | GGTGAGGAGG AGCCCTG<br>AGGTACAAAG GCCCCGG<br>AACAGGCATT TCCCC                       |                |                          |                          |  |                   |                                  |           |                 |           |  |
| >gnl dt                   |  |                |                          |                          |  |                   |                                  |           |                 |           |  |
| CAGC.<br>RACT.            | Submitted Freque   | ncy for a      | s38342                   | 908                      |  |                   |                                  | t         |                 |           |  |
| CCC.b<br>cttm             |  |                |                          |                          |  |                   |                                  |           |                 |           |  |
| ccap<br>GTTD<br>S<br>CTCT | Population ID<br>-Class  | Sample<br>(2n) | Major<br>Allele<br>Freq. | Minor<br>Allele<br>Freq. | Estimated<br>Heterozygosity<br>+/-std.err. | Genotype<br>Freq. | Submitted<br>Hetero-<br>zygosity |           | nission<br>atch | Submitter |  |
| TTAC'                     | EGP AD-PANEL   | 28             | G=0.857                  | C=0.143                  |  |                   |                                  | NOS2A-EGP | AD-051605       | EGP SNPS  |  |
| TGTG'<br>CTGA             | EGP ASIAN-PANEL  | 44             | G=0.818                  | C=0.182                  |  |                   |                                  | NOS2A-EGP | ASIAN-051605    | EGP SNPS  |  |
|                           | EGP YORUB-PANEL  | 24             | G=0.875                  | C=0.125                  |  |                   |                                  | NOS2A-EGP | YORUB-051605    | EGP SNPS  |  |
|                           | EGP CEPH-PANEL   | 44             | G=0.955                  | C=0.045                  |  |                   |                                  | NOS2A-EGP | CEPH-051605     | EGP SNPS  |  |

| S NCBI Entrez Gene   | My NCB  |           |
|--|---|-----------|
| Alton         Nature         Fallen         Fallen </td <td>Links<br/>Conserved Domains<br/>Genome<br/>GEO Profiles<br/>Hemolo Gene<br/>Map Viewer<br/>Nucleotide</td> <td>]</td>   | Links<br>Conserved Domains<br>Genome<br>GEO Profiles<br>Hemolo Gene<br>Map Viewer<br>Nucleotide | ]         |
| T1: NOS2A minite exide synthase 2.A (inducible, hepatocytes) [Homo sapiens]     GaseDL 4843 Finany source <u>HORMO7873</u> Summary   | OMIM<br>Full text in PMC<br>Probe<br>Protein  | 06        |
| Official Symbol. 2002L, and Yours: etc. and exprained 2A (deckeds), ispanoytes) perioded by <u>UDDO Ones</u><br>for the related <u>UDDO Ones</u> (MARCH 1997)<br>(2004) (200 | HPRD<br>KEGG<br>ModelMaker  | ains<br>1 |
| Genomic regions, transcripts, and products   | PharmOVB<br>UCSC<br>Universe  |           |
| MC_000017         40007001           9         4000000000000000000000000000000000000   | Entrez Gene Info     Feedback   |           |

| s s   |                                | 2  | 1                           |   |  |
|---|--------------------------------|--|-----------------------------|---|--|
| Search St   | P Receiption Protein Genome St | Gothie Poceet Tablony  | WWW BOOKS DWY               |   |  |
| RULD  |                                | E GENOTYPE REPORT  |                             |   |  |
| Contact Un<br>de SP7 Housey<br>SP7 Science Pr                   |                                | iene ID Gene Symbol  | Che Report                  |   |  |
| Automoted door Surgery  | 2010/01/02/02/02               | 4343 NOS2A<br>time (LD) Analysis (click on al                      | 17 http://www.amiliania.com | 2 |  |
| FTP Doveland<br>Server<br>Overlag Sheri<br>Redd Tablery         |                                | Solumiter Handle <u>CSHL HAR</u><br>Population ID: <u>HAPMAP</u>   |                             |   |  |
| Fault Prese   |                                | Submitter Handle CSHL-HAD  | MAP                         |   |  |
| DOCUMENT<br>Endl Release<br>7A/Q<br>Uncert                      | ATION                          | Population ID: HAPMAP  |                             |   |  |
| South Path<br>doll P House<br>Charles<br>Eline In Salard        |                                | Subastrer Handle <u>CSHL-HAI</u><br>Population ID: <u>HAPMAP</u> . |                             |   |  |
| Aut202 Summ<br>Setem<br>Dataset<br>FOR<br>Distance              |                                | Submitter Handle <u>CSHL HAR</u><br>Population ID: <u>HAPMAP</u>   | RI I                        |   |  |
| United<br>Charges E<br>Oncorps<br>Data Formation<br>Printmation |                                | Submitter Handle: PERLECE<br>Population ID: <u>APD APR</u>         |                             |   |  |
| Cooperation<br>SEARCH<br>Robert 100<br>Dista AR                 |                                | Solution Handle PERLIGE<br>Population ID: APD CHN                  |                             |   |  |
| Dow An<br>Bart 207<br>Beth Corry                                |                                | Submitter Handle PERLEGE   |                             |   |  |

| CENT CONTRACT AND AND A TRUE TREADING TREADING TREADING TREADING     CENT THE CONTRACT OF CONTRAC  | IbSNP Ge     | notvne a     | nd Allele Free | nuency | Report   | Build 12 | 25 |
|--|--------------|--------------|----------------|--------|----------|----------|----|
| Conserve 443 ECCA 13 mil mo<br>(Conserve on Harman Marcanadra Statistics)<br>FEELEDAAL MARCHART END Parts.<br>SEELEDAAL AND END PARTS.<br>SEELEDAAL MARCHART END Parts.   |              | ENE          |                |        |          |          |    |
| Generges en Henne Henn genergen Subject<br>#ERLEGEN.ATD. TUR FANEL Par Dess.<br>#ERLEGEN.ATD. CHU FANEL Par Dess.<br>#ERLEGEN.ATD. CHU FANEL Par Dess.<br>CSHL HAPAAL HENME-SUP Par Dess.<br>CSHL HAPAAL HENME-SUP Far Dess.<br>CSHL HAPAAL HENME-SUP Far Dess.<br>CSHL HAPAAL HENME-SUP Far Dess.<br>CSHL HAPAAL HENME-SUP Far Dess.<br>#EP SUPS-ESF CHUF-FANEL Par Dess.<br>#EP SUPS-ESF CHUF-FANEL Par Dess.<br>#EP SUPS-ESF CHUF-FANEL Par Dess.   |              |              |                |        |          |          |    |
| SERIEMAN ATE EUT PANE Paches.<br>SERIEMAN ATE EUT PANE Paches.<br>SERIEMATE ATE ANNE Paches.<br>SERIE ADVACH Bandae CEU Paches.<br>SERIE ADVACH SERIEMAE THE Paches.<br>SERIEMAE THE SERIEMAE THE PACHES.<br>SERIEMAE THE PACH |              |              |                | 17     | xml gzip |          |    |
| VERLEGELAFE AFE LANE En Dem<br>PERLEGELAFE AFE LANE En Dem<br>CERLEGELAFE CHI FAHE Par Dem<br>CERLEGELAFE (SAME-SEE Par Dem<br>CERLEGELAFE SAME-SEE Par Dem<br>CERLEGELAFE AMME-STER Par Dem<br>CERLEGELAFE AMME-STER Par Dem<br>CERLEGELAFE AMME-STER Par Dem<br>SEP SITE-SEF LIES PARE Par Dem<br>SEP SITE-SEF LIES PARE Par Dem<br>SEP SITE-SEF LIES PARE Par Dem   |              |              |                |        |          |          |    |
| WEILIGHAATD CHD FAND, Ion Inn.<br>WEILIGHAAF MAMA-DY Ion Don.<br>WIILIGHAAF MAMA-DY Ion Don.<br>WIILIGHAAF MAMA-DY Ion Ion.<br>WIILIGHAAF MAMA-DY Ion Ion.<br>WIILIGHAAF MAMA-DY Ion Ion.<br>WIILIGHAAF MAMA-DY Ion Ion.<br>WIII SANAAF MAMA-DY Ion Ion.<br>WIII SANAAF MAMA-DY Ion Ion.<br>WIII SANAAF MAMA-DY Ion Ion.<br>WIII SANAAF MAMA-DY Ion.<br>WIII SANAAF MAMA-DY Ion.   |              |              |                |        |          |          |    |
| CEIL BAVAR Hawke-CEI Pa Dan.<br>(SEIL BAVAR Hawke-FI Pa Dan.<br>CEIL BAVAR Hawke-FI Pa Dan.<br>CEIL BAVAR Hawke-FI Pa Dan.<br>CEIL BAVAR Hawke-FI Pa Dan.<br>ECI BAVE-FI VOIT-FAUL Pa Dan.<br>ECI SUP-ECI CEIP-FAUL Pa Dan.<br>ECI SUP-ECI CEIP-FAUL Pa Dan.<br>ECI SUP-ECI CEIP-FAUL Pa Dan.  |              |              |                |        |          |          |    |
| CSHL HAPMAP HapMap-TF Pop Den.<br>CSHL HAPMAP HapMap-TF Pop Den.<br>GST SURV.SET, VINTER-ANGLI For Den.<br>GST SURV-SET HIST-PANIL Pop Den.<br>GST SURV-SET CEPH-PANIL Pop Den.<br>GST SURV-SET AD-PANIL Pop Den.  |              |              |                |        |          |          |    |
| CETLE ADVIADA Harving, The Dates<br>GRE ATTESECT YOND-FAIGU For Dates<br>GRE ATTESECT YOND-FAIGU For Dates<br>GRE ATTESECT CETH-FAIGU For Dates<br>GRE ATTESECT CETH-FAIGU For Dates.  | CSHL-HAPMA   | P-HapMap-HCI | B Pop Desc.    |        |          |          |    |
| EGF_SNPS-EGF_YOKUE-PANEL Pop Desc.<br>EGF_SNPS-EGF_USE-PANEL Pop Desc.<br>EGF_SNPS-EGF_CEFFANEL Pop Desc.<br>EGF_SNPS-EGF_AD-PANEL Pop Desc.   | CSHL-HAPMA   | P-HapMap-JPT | Pop Desc.      |        |          |          |    |
| EGP SNPS-EGP HISP-PANEL Pop Desc.<br>EGP SNPS-EGP CEPH-PANEL Pop Desc.<br>EGP SNPS-EGP AD-PANEL Pop Desc.  | •CSHL-HAPMA  | P-HapMap-YRI | Pop Desc.      |        |          |          |    |
| EGF SNPS-EGF CEPH-PANEL Pop Desc.<br>EGF SNPS-EGF AD-PANEL, Pop Desc.  | EGP SNPS-EG  | P YORUB-PAI  | NEL Pop Desc.  |        |          |          |    |
| EGP_SNPS-EGP_AD-PANEL Pop Desc.  | EGP SNPS-EG  | P HISP-PANEL | L Pop Desc.    |        |          |          |    |
|  | •EGP SNPS-EG | P CEPH-PANE  | L Pop Desc.    |        |          |          |    |
| EGP_SNPS-EGP_ASIAN-PANEL Pop Desc.   | EGP_SNPS-EG  | P AD-PANEL   | Pop Desc.      |        |          |          |    |
|  | EGP SNPS-EG  | P ASIAN-PAN  | EL Pop Desc.   |        |          |          |    |

| A GE   |  |  |  | ,                                       |  | t <u>Build1</u>  | 20   |   |   |   |  |   |  |
|--|--|--|--|---|--|--|--|---|---|---|--|---|--|
|  | INE  |  |  |   | +7   |  |  |   |   |   |  |   |  |
| 110  | GENOTYP  | E REPORT   |  |   | -  |  |  |   |   |   |  |   |  |
| Organism   | Gens ID  | Gene Symbo   |  | hr                                      | Report   |  |  |   |   |   |  |   |  |
| Homo azesene   | 4843   | NOS2A  |  | 17                                      | and gain   |  |  |   |   |   |  |   |  |
| Genotypes on H   | inn in <u>Konse ops</u> i  | Subjects   |  |   |  |  |  |   |   |   |  |   |  |
| ·PERLEGEN-AF   | D EUR PANEL  | Pop Derc.  |  |   |  |  |  |   |   |   |  |   |  |
| *FERLEGEN-AF   | D AFR PANEL  | Pop Desc.  |  |   |  |  |  |   |   |   |  |   |  |
| *PERLECEN-AF   | D CHN PANEL  | Pop Derr.  |  |   |  |  |  |   |   |   |  |   |  |
| CSHL-HAPMAN  | P-HapMag-CEU   | Pop Desc.  |  |   |  |  |  |   |   |   |  |   |  |
| CSHL-HAPMAN  | P-HapMap-HCB   | Pep Desc.  |  |   |  |  |  |   |   |   |  |   |  |
| ·CSHL-HAPMAI   | P-HapMap-JPT   | lop-Dest.  |  |   |  |  |  |   |   |   |  |   |  |
| CSHL-HAPMAN  | P-HaoMao-YRI   | Post Desc.   |  |   |  |  |  |   |   |   |  |   |  |
| -EOP SNPS-EOR  | P YORUB-PAN  | EL Pop Desc.   |  |   |  |  |  |   |   |   |  |   |  |
| Individual Group   | di SNP Pedig   | ee Individual Se   | s Mother                                       | Father                                  | 1528944207   | 1128942368   | 1128993833   | 116966563                                     | 1523998832                                    | 1423998833                                    | 1413998330   | 08072199  |  |
| Sub-Saharan Africa.  | 4 43   | 61 NA18502 F   | 0  | 0                                       | GG   | DI   | CIC  | AIG   |   | GIG   | 8/8  | G/G   | G/G  |
|  |  |  |  |   |  |  |  |   |   |   |  |   |  |
|  | a <u>43</u>  | 62 NA18504 M   |  | 0                                       |  | TYT  | eic .  | A/A   | cic   | 0/0   |  |   | G/G  |
| Sub-Saharan Africa   | a <u>43</u><br>a <u>43</u>   | 62 NA18504 M<br>71 NA18870 F   | 0  | 0                                       | GIG  | TUT  | CIC  | AVG   | C/C   | G/G   | A/A  | G/G   | G/G  |
| Sub-Saharan Africa<br>Sub-Saharan Africa   | 43<br>43<br>43   | 62 NA18504 M<br>71 NA18870 F<br>74 NA18855 F   | 0  | 0                                       | 013<br>013   | DT<br>DT   | aic<br>aic   | DIA<br>AVA                                    | c/c   | 0/0<br>G/0                                    | A/A<br>A/A   | a.a<br>a.a  | 6.6<br>6.6   |
| Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa   | 43<br>43<br>43<br>43   | 62 NA18304 M<br>71 NA18870 F<br>74 NA18855 F<br>88 NA19137 F   | 0<br>0<br>0                                    | 0<br>0<br>0                             | 013<br>013<br>015  | DI<br>DI<br>DI   | aic<br>aic<br>aic                                  | A/G<br>A/A<br>A/A                             | cic<br>cic                                    | 0/0<br>G/G<br>G/G                             | A/A<br>A/A<br>A/A  | 6 6<br>6 6<br>6 6   | 6.6<br>6.6<br>6.6  |
| Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa   | 43<br>43<br>43<br>43<br>43<br>43<br>43                                       | 62 NA18504 M<br>71 NA18870 F<br>74 NA18875 F<br>88 NA19137 F<br>90 NA19200 M   | 0 0 0 0 0                                      | 0 0 0 0 0                               | 013<br>013<br>015<br>015                                 | LU<br>LU<br>LU<br>LU   | aic<br>aic<br>aic<br>aic                           | A/G<br>A/A<br>A/A<br>A/G                      | cic<br>cic<br>cic                             | 0/0<br>G/G<br>G/G<br>G/G                      | A/A<br>A/A<br>A/A<br>A/A   | 6 6<br>6 6<br>6 6<br>6 6  | G.G<br>G.G<br>G.G<br>G.G   |
| Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa   | a 43<br>a 43<br>a 43<br>a 43<br>a 43<br>a 43<br>a 43<br>a 43                 | 62 NA18304 M<br>71 NA18370 F<br>74 NA18355 F<br>88 NA19137 F<br>90 NA19330 M<br>90 NA19331 F   | 0 0 0 0 0 0 0 0                                | 0<br>0<br>0<br>0                        | 8/3<br>9/3<br>9/3<br>8/3                                 | DA<br>DA<br>DA<br>DA<br>DA<br>DA                                   | ac<br>ac<br>ac<br>ac                               | DIA<br>AIA<br>AIA<br>DIA<br>AIA               | 40<br>00<br>00<br>00<br>00                    | 0/0<br>G/G<br>G/G<br>G/G<br>0/0               | A/A<br>A/A<br>A/A<br>A/A<br>A/A                                      | G/G<br>G/G<br>G/G<br>G/G<br>G/G   | G/G<br>G/G<br>G/G<br>G/G<br>G/T                                    |
| Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa   | a 43<br>a 43<br>a 43<br>a 43<br>a 43<br>a 43<br>a 43<br>a 43                 | 62 NA18504 M<br>71 NA18370 F<br>74 NA18375 F<br>88 NA19137 F<br>90 NA19201 M<br>90 NA19201 F<br>92 NA19203 M   | 0        | 0<br>0<br>0<br>0<br>0<br>0              | 0/3<br>0/3<br>0/3<br>0/3<br>0/3                          | 101<br>101<br>101<br>101<br>101<br>101                             | aic<br>aic<br>aic<br>aic<br>aic                    | DIA<br>AIA<br>AIA<br>DIA<br>AIA<br>DIA<br>DIA | 00<br>00<br>00<br>00<br>00<br>00<br>00        | 0/0<br>G/G<br>G/G<br>G/G<br>Q/G<br>Q/G        | A/A<br>A/A<br>A/A<br>A/A<br>A/A<br>A/A                               | G/G<br>G/G<br>G/G<br>G/G<br>G/G<br>A/G                                    | G.G<br>G.G<br>G.G<br>G.T<br>G.G                                    |
| Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa                       | a 43<br>a 43<br>a 43<br>a 43<br>a 43<br>a 43<br>a 43<br>a 43                 | KI         NA18504         M           71         NA18870         F           74         NA18255         F           88         NA19137         F           90         NA19201         M           90         NA19201         F           92         NA19201         F           92         NA19201         M           93         NA19201         M           94         NA19201         M  | 0        | 0 | 0/3<br>0/3<br>0/3<br>0/3<br>0/3<br>0/3<br>0/3<br>0/3     | 101<br>101<br>101<br>101<br>101<br>101<br>101                      | 80<br>00<br>00<br>00<br>00<br>00<br>00<br>00<br>00 | A/G<br>A/A<br>A/G<br>A/G<br>A/G               | 010<br>010<br>010<br>010<br>010<br>010        | 0/0<br>0/0<br>0/0<br>0/0<br>0/0               | A/A<br>A/A<br>A/A<br>A/A<br>A/A<br>A/A<br>A/A                        | 6.6<br>6.6<br>6.6<br>6.6<br>6.6<br>6.6<br>8.6<br>8.6<br>8.6               | 6.6<br>6.6<br>6.6<br>6.6<br>6.7<br>6.7<br>6.6<br>6.7               |
| Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa | a 43<br>a 43<br>a 43<br>a 42<br>a 42<br>a 43<br>a 43<br>a 43<br>a 43<br>a 44 | NA19504         M           71         NA18370         F           74         NA18355         F           75         NA19137         F           90         NA19200         M           90         NA19201         F           92         NA19203         M           93         NA19203         M           94         NA19203         M           95         NA19203         M           96         NA19203         M           97         NA19203         M           98         NA19233         M           90         NA19233         M | 0<br>0<br>0<br>0<br>0<br>0<br>0<br>0<br>0      | 0<br>0<br>0<br>0<br>0<br>0              | 00<br>00<br>00<br>00<br>00<br>00<br>00<br>00<br>00       | 101<br>101<br>101<br>101<br>101<br>101<br>101<br>101<br>101<br>101 | 80<br>00<br>00<br>00<br>00<br>00<br>00<br>00<br>00 | A/G<br>A/A<br>A/G<br>A/G<br>A/G<br>A/G<br>A/A | 010<br>010<br>010<br>010<br>010<br>010<br>010 | 0/0<br>G/G<br>G/G<br>G/G<br>G/G<br>G/G<br>G/G | Х/А<br>Х/А<br>Х/А<br>Х/А<br>Х/А<br>Х/А<br>Х/А<br>Х/А                 | 6.6<br>6.6<br>6.6<br>6.6<br>6.6<br>4.6<br>6.6<br>4.6<br>4.6<br>4.6        | 6.6<br>6.6<br>6.6<br>6.7<br>6.7<br>6.7<br>6.6<br>6.7<br>6.6<br>6.6 |
| Sulo-Saharan Adira<br>Sulo-Saharan Adira<br>Sulo-Saharan Adira<br>Sulo-Saharan Adira<br>Sulo-Saharan Adira<br>Sulo-Saharan Adira<br>Sulo-Saharan Adira<br>Sulo-Saharan Adira | a 43<br>a 43<br>a 43<br>a 43<br>a 43<br>a 43<br>a 43<br>a 43                 | 62 NA18504 M<br>71 NA18370 F<br>74 NA18375 F<br>88 NA19137 F<br>90 NA19201 M<br>90 NA19201 F<br>92 NA19203 M   | 0        | 0<br>0<br>0<br>0<br>0<br>0              | 0/3<br>0/3<br>0/3<br>0/3<br>0/3                          | 101<br>101<br>101<br>101<br>101<br>101                             | aic<br>aic<br>aic<br>aic<br>aic                    | DIA<br>AIA<br>AIA<br>DIA<br>AIA<br>DIA<br>DIA | 00<br>00<br>00<br>00<br>00<br>00<br>00        | 0/0<br>G/G<br>G/G<br>G/G<br>Q/G<br>Q/G        | A/A<br>A/A<br>A/A<br>A/A<br>A/A<br>A/A                               | G/G<br>G/G<br>G/G<br>G/G<br>G/G<br>A/G                                    | 6.6<br>6.6<br>6.6<br>6.7<br>6.7<br>6.7                             |
| Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa<br>Sub-Saharan Africa   | 43<br>43<br>43<br>44<br>44<br>44<br>44<br>44<br>44<br>44<br>44               | KI         NA18504         M           71         NA18870         F           74         NA18255         F           88         NA19137         F           90         NA19201         M           90         NA19201         F           92         NA19201         F           92         NA19201         M           93         NA19201         M           94         NA19201         M  | 0<br>0<br>0<br>0<br>0<br>0<br>0<br>0<br>0<br>0 | 0<br>0<br>0<br>0<br>0<br>0<br>0         | 96<br>96<br>96<br>96<br>96<br>96<br>96<br>96<br>96<br>96 | 101<br>101<br>101<br>101<br>101<br>101<br>101                      | 80<br>00<br>00<br>00<br>00<br>00<br>00<br>00       | A/G<br>A/A<br>A/G<br>A/G<br>A/G               |   | 0/0<br>0/0<br>0/0<br>0/0<br>0/0               | NA<br>NA<br>NA<br>NA<br>NA<br>NA<br>NA<br>NA<br>NA<br>NA<br>NA<br>NA | 6.6<br>6.6<br>6.6<br>6.6<br>6.6<br>4.6<br>6.6<br>4.6<br>6.6<br>4.6<br>6.6 | 6.6<br>6.6<br>6.6<br>6.6<br>6.7<br>6.7<br>6.6<br>6.7               |

| Minimal SNP information  | n for genotyping/characterization    |
|--|--------------------------------------|
| <ul> <li>What is the SNP? Flankin</li> <li>✓ FASTA format</li> </ul>         | ng sequence and alleles.             |
| >snp_name<br>ACCGAGTAGCCAG<br>[A/G]<br>ACTGGGATAGAAC                         | dbSNP - data is there                |
| dbSNP reference SNP # (  | rs #)                                |
| <ul><li>✓ picture of gene with ma</li><li>How was it discovered? M</li></ul> | 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!

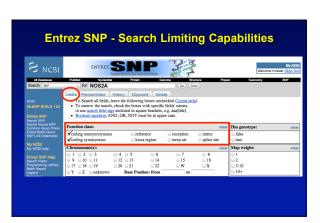
| S NCBI Entrez Gene  | · · · · · · · · · · · · · · · · · · ·                | My NOR                       | u<br>Line e |
|---|--|------------------------------|-------------|
| All Databases Robling Book  | in Oceanie Blockie PMC                               | Table Of Contents      Links |             |
| Limits PreviewIndex History Clipboard Details   |  | Conserved Domains            | -           |
| Dealer Full Perport   |  | Genome<br>GED Profiles       | 12          |
|   |  | HomoloGene                   |             |
| All: 1 Genes Genomes: 1 SNP Geneview: 1 🛠   |  | Map Viewer<br>Nucleatide     | 18          |
| □ 1: NOS2A nitric oxide synthase 2A (inducible, he<br>GeneID: 4843 Frimery source HONC 7873 | patocytes) [Hamo sapiens]                            | OMIM                         | 06          |
|   |  | Full text in PMC<br>Probe    |             |
| Summary   |  | Protein                      |             |
| Official Symbol: NOS2A and Name: nitric oxide synthase 2A (ir                               | ducible, hepatocytes) provided by <u>HUGO Gene</u>   | PubMed                       |             |
| See related: HPRD 01225, MIM:163730<br>Gene type: protein podra                             |  | SNP                          |             |
| Gene name: NOS2A  | -  | SNP: Genotype                |             |
| Gene description: nitric oxide synthase 2A (inducible, hepatocytes                          | 0  | SNP: GeneView                |             |
| RefSeq status: Reviewed   |  | UniSTS                       |             |
| Organium: Homo aupiana  |  | AceView                      |             |
| Line+ge: Bukaryota; Metazoa; Chordata; Craniata; Vertebrat<br>Hominidae; Homo               | a; Buteleostomi; Mammalia; Butheria; Buarch          | CCDS<br>Ensembl              |             |
| Gene aliases: NOS. INOS. NOS2; HEP-NOS  |  | Ensembl<br>Exidence Viewer   |             |
| Summary: Nitric oxide is a reactive free radical which acts as a bio                        | logic mediator in reveral processes, including resp  | ODB                          |             |
| antitumoral activities. This gene encodes a nitric oxide synthase which                     | h is expressed in liver and is inducible by a combi- | HONC                         | én.         |
| cytokines. Three related pseudogenes are located within the Smith-                          | Magenis syndrome region on chromosome 17. Alt        | HPRD<br>KEGG                 |             |
| two transcript variants encoding different isoforms.  |  | ModelMaker                   |             |
| Genomic regions, transcripts, and products  |  | Pharmól/B                    |             |
| (minus strand) RefSeg below   |  | VCSC                         |             |
| NC.0H017  |  | LinkOut                      |             |

| Entrez  | SNP - direct dbSNP querying   |
|---|---|
| S NCBI  | ENTREZ SNPP   |
| Al Information Public Search, Step  | nd Abdatide Attent Genome Stuckte Dropet Texanomy SHP<br>■ Bor<br>Go Goer   |
| NCB<br>d5SNP BUILD 125  | Lints Previewtindex Hobory Clipboard Details Deploy Graph Store 20 Storby Send to a   |
| Entres SIP  | All: 334 Human: 334 Mouse: 0 NEW: 170 Other Organisms: 0 UPDATE 85 🛠<br>Rems 1 - 20 of 334 Page 1 of 17 Next  |
| Basich Magise SNP<br>Common Guary Filtern<br>Entraz Black Guary<br>SNP Link Datamodel | Ling8070472 (Rees appens)     Units     Takalakatar to choose ( 2/9) To contract to the  |
| My NCBI<br>My NCBI keb  | Image: Control                    |
| Entrez SNP Help<br>Seaschable FAG<br>Blaach Fielda<br>Programming Utilities           |   |
| Batch Preport<br>Legend<br>Bramales<br>straWP Home Place                              | Str traditional (Renew Suptem)     Units     Matching Control (V/C) associated in the second se |
| Overview<br>Entrez Help   | Martine         Converse         No. Converse  |
| General help<br>Livits<br>Provinsioninges<br>History                                  |   |
| Classore<br>Debils<br>Other Services  | El Se <u>re22944100</u> //Koreo signien/  |
| Japmal Browser<br>Mc6H Browser<br>Claston Mather<br>Cleation Mather                   |   |
|   | □ 6: <u>#28999376</u> (Horeo septens) UH6   |

| S NCBI   | ENTREZ SA  | Welcome mieder. (Sign. Out   |
|--|--|--|
| Search Search  | <ul> <li>Graphic Summary<br/>Brief</li> <li>Flat File</li> </ul> | Concres Statute Ayest Tecrosy Stat<br>Go Cear<br>Clipboard Details |
| NCBI<br>dbSNP BUILD 124  | ASN.1<br>XML   | Send to  |
| Entroz SNP<br>Search SNP<br>Search Nocae SNP<br>Commo Query Filens             | FASTA<br>RS CLuster Report<br>SS Exemplar List                   | File Page 1 of 5 Next<br>Clipboard Line                            |
| Entrez Betch Query<br>SNP Link Determodel<br>My NCBI                           | Chromosome Report<br>Genotype                                    | C JAGACOCTITITITITICTICAOTGAA                                      |
| Mỹ NOBI help<br>Entrez SNP Help<br>Search Fields                               | Genotype Detail<br>Genotype XML                                  | Link   |
| Programming Utilities<br>Backh Report<br>Legend<br>Examples<br>ch5NP Nome Pase | Summary<br>dbSNP Batch Report<br>UI List                         | in 20 No. SHOW (   |
| Costor Hote Hage<br>Overview<br>Entrez Help<br>General Help                    | Gbench Export<br>LinkOut   | /7) TOTO DE ANTOCTO COTO CADERACE                                  |
| Canversa help<br>Limita<br>Proviewindox<br>History<br>Clichoant                | Gene Genotype Links<br>Gene Links<br>HomoloGene Links            | Line   |
| Other Services   | Nucleotide Links<br>OMIM Links                                   | 10 TO NO CAN THE TOP OF TO     |
| MisSel Browser<br>Citation Matcher<br>Clinical Queries                         | PMC Links<br>Protein Links                                       | SNPSCOCANCER   |
|  | PubMed Links<br>Structure Links<br>Taxonomy Links                | (s 30 - ML (SHM - (SHM - SH) V C                                   |
|  | UniGene Links<br>UniSTS Links                                    |  |

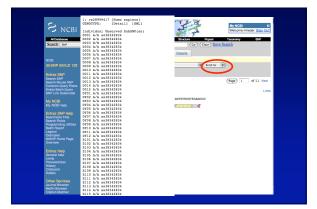
| Er  | ntrez SNP - Parseable Multi-SNP reports  | ;  |
|---|--|----|
| Allow   |  | 2  |
| Common<br>Entrez Bat  | СТОСТВИИТО ОТАЛСТВОВО СОСОЛЖЕТТ СООТОРАЛИТ НАТАЛЛИВИ СОСТОТОТО ТОТОТАЛЛО ТИКТОССАЛ<br>ИНСТИЛИКА СОЛЖИТИК ОБЛИКАТТ ТОТТОТТОТ<br>К<br>ВОЛЖИЛИСТ АНТОГОЛЬВИ ПОЛУТИТИТИТИ<br>СООТОЛИЦИИ ПОЛУТИТИТИТИ ОПИСАТИ СОСОЛЮТКА ТАСТОТОТОТ ИНТЕТНОВА ПОВОЛОГОЛА<br>САТОЛИТИКА ПАТОЛИСТИК ТОТОСЛАВИТ   | 1  |
| Programm<br>Batch Rep<br>Legend   | 3: r235729 [Bee spice]<br>spice] (Addy [T237]) [Licker_content]] (add/9406] (Spic]as-1[Licker_777 [no]-speced() [bald-108<br>spic](Addy [T237]) [Licker_content]] (add/9406] (Spic](add/9406] (Add/9406] (Add/9406) [bald-108<br>spic](Add/9406) [Add/9406] [Add/9406] (Add/9406) [Add/9406] (Add/9406) [Add/9406] [Add/94   | TE |
| Examples<br>dbSNP Ho<br>Overview<br>Entrez Hi<br>General hi<br>Limits<br>Previewdin<br>History<br>Clipboard | 4: relisti [Bee spice]<br>6: relisti [Bee spice]<br>relisti [Bee spice]<br>rel | TE |

| Entrez SNP - Parseable Multi-SNP reports  |  |  |  |  |
|---|--|--|--|--|
| i 1999/1 (Box equar)<br>1999/1 (Box equar)<br>1999  |  |  |  |  |
| 2: rdf01046 [Emon sapins]<br>42: rdf01046 [Emon sapins]<br>42: rdf01046 [Emon [emony=00] rabatterilar-80 [updated 15/4/2004 15:21:00<br>42: rdf0105 [Emony=01] rdf1 [update]<br>40: usidated=00 [Emony=01] [update]<br>40: usidated=00 [update]  |  |  |  |  |
| ], raj37378 [Home empired]<br>201723728 [Home empired]<br>201723728 [Home empired]<br>201723728 [Home empired]<br>2017247 [Home, Home, Home   |  |  |  |  |
| Journel Beauwr<br>Calculur gawar<br>Calculur gawar<br>Calculur gawar<br>Unifer Charles<br>Dawar<br>Calculur gawar<br>Calculur gawar<br>Calcul |  |  |  |  |



| Construction         Construction<  |  | E      | ntrez SNP                  | - Search Limits  |                             |
|---|--|--------|----------------------------|--|-----------------------------|
| Statts         10 (f) rests         (f) rests         (f) rests           And<br>distilitie         Umits         Prevendinder         Hationy         Cippoard         Datalis           And<br>distilitie         Umits         Prevendinder         Hationy         Die of<br>Prevendinder         Fill           And<br>Bref         Bref         Fill         Fill         Fill         Fill         Fill           Anno         Bref         Fill  |  |        | Single Nucleoti            | Stat New York of the State of t | Welcome mrieder. [Sign Out] |
| Note<br>Server Sub<br>Server Server Se |  |        |                            |  |                             |
| address DAD     Direct     Sort B     Sort B     Sort B       Brend SAD     Field     Field     Field       Brend SAD     Field     Field     Field       ASN 1     ASN 1     Field     Field       Brend SAD     Field     Field     Field       Brend SAD     Field     Field     Field       Chromoson SAD     Field     Field     Field       Som San SAD     Field     Field     Field       Chromoson San San San San San San San San San Sa  |  | เฮน    | mits Preview/Index Hist    | ory Clipboard Details  |                             |
| address DAD     Direct     Sort B     Sort B     Sort B       Brend SAD     Field     Field     Field       Brend SAD     Field     Field     Field       ASN 1     ASN 1     Field     Field       Brend SAD     Field     Field     Field       Brend SAD     Field     Field     Field       Chromoson SAD     Field     Field     Field       Som San SAD     Field     Field     Field       Chromoson San San San San San San San San San Sa  |  | Limits | coding nonsynon, coding sy | nonymous   |                             |
| Entre Sub<br>Sector Academy File         File         File         File           File         File         File         File         File           AVN 12         XML         File         File         File           SW Link Lawrender         In         XML         File         File           SW Link Lawrender         In         XML         File         File           SW Link Lawrender         In         Stater Report         File         File           SW N2BL         Stater Report         Chromosome Report         Chromosome Report         Concrupte Data           Controls on Proportional Concord Con   |  |        |                            |  |                             |
| Entry         Direction         File         File           Variat         File         File  |  | _      |                            |  |                             |
| Same in Subset Burg<br>South Law South Law Sou  |  | AI:    |                            |  |                             |
| Control Contro Control Conte Control Control Control Control Control Control Co   |  |        |                            | Clipboard  | age 1 of 2 Next             |
| Story Issue Stateworks         IF         FACTA         IA/S (Enter Report<br>SE Schwarks Report<br>SE Schwarks Report<br>Chromosoms  | Common Query Filters                     |        |                            | Chipboard  |                             |
| Ny NCB1         RS CLuster Report         I/V0[100000000000000000000000000000000000   | Entrez Batch Query<br>SNP Link Datamodel | ⊟1:    |                            |  | Links                       |
| W/ ADDI         Stampler List         Interview 30% all           Second 20% all         Chromosome Report         Lit           Chromosome Report         Carolype         Lit           Second 20% all         Carolype         Lit           Carolype         Second 20% all         Lit           Carolype         Lit         Lit         Lit           Carolype         Lit         Lit         Lit           Carolype         Lit         Lit         Lit           Carolype         Lit         Lit         Lit <td< td=""><td></td><td></td><td></td><td>(b/G)CTOTTGOTGOTGACCAOTACOTTTG</td><td></td></td<>  |  |        |                            | (b/G)CTOTTGOTGOTGACCAOTACOTTTG   |                             |
| Construction     Control Source Report     Control Source Report     Link       Control Source Report     Control Source Report     Control Source Report     Link       Source Report Report     Control Source Report     Control Source Report     Link       Link     Control Source Report     Control Source Report     Link       Control Source Report     Control Source Report     Link       Control Source Report     Link     Link     Link       Control Source Report     Link     Control Source Report     Link       Control Source Report     Control Source Report     Link  |  |        |                            | []   |                             |
| Sama bio span and spa  |  |        |                            | No 3D No CMM   |                             |
| Sample and USA<br>Sample and USA  |  | □ 2:   |                            |  | Links                       |
| Programmer         Construction         Construction           Summary         Construction         Construction           Construction         Construction         Construction  |  |        |                            |  |                             |
| Ligner<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Support<br>Supp  | Programming Utilities                    |        |                            | [6/7]CTCCATAGTTTCCAGGTATGTCTG6   |                             |
| Defenses         3-4         disSP Batch Report         LL           Util tit         Util tit         LL/10[TYT00A0CA0CCA0CAACAACCTTC         LL           Centre Nego         LL/10[TYT00A0CA0CCA0CAACAACCTTC         LL           Centre Nego         LL         Centre Nego         LL           LunkOut         LL/10[TYT00A0CA0CCA0CAACAACCTTC         LL           Centre Nego         Centre Centryse LInks         Centre Centryse LInks         LL           Centre Nego         LL         Centre Centryse LInks         LL           Centre Nego         LL         Centre Centryse LInks         LL           Centre Nego         Centre Centryse LInks         LL           Centre Nego         Centre Centryse LInks         LL           Centre Centryse LInks         Centre Centryse LInks         LL           Centre Nego         LL         Centre Centryse LInks         LL   | Batch Report                             |        |                            | No 3D No CMM Comments V G  |                             |
| Converse         UL 14         Call Participation         Call Partin Parited Participation         Call Partin Pa  | Examples                                 | - 2.   |                            |  | Links                       |
| Entre Hego<br>Gaussian Multip         U/L13*         [L/o] TPTGGGLAGGCAGGLAGGCAGCTAGCTAGCTAGC<br>Gaussian Multip           Caussian Multip         LinkOut         Terrore Gaussian Multip         LinkOut           Cence Genotype Links<br>relaxy<br>relaxy<br>Gaussian Multip         Cence Genotype Links<br>Cence Genotype Links         LinkOut         LinkOut           Cence Genotype Links<br>HomoloCore Links         Cence Genotype Links         LinkOut         LinkOut   |  | - 3.   |                            |  | Cirks                       |
| Canada Nava<br>LinkOut  LinkOut   |  |        |                            | [A/G]TTTGGAGCAGGCCAAGGAAGCCTTC   |                             |
| Linis<br>Menovembede<br>Hausy<br>Concert Links<br>Concert Links<br>(c/*) anoccastorococcachochtec   |  |        |                            | No 2D No CMM   |                             |
| Perviewindex Hittory Clipboard Clipb  |  |        |                            |  |                             |
| Clipboard HomoloGene Links (C/9)0A00C0A0T07000CCACA0CATCC   |  | ⊟4:    |                            |  | Links                       |
| HomoloGene Links  |  |        |                            | CC/710A0000A0T070000ChCA00ATCC   |                             |
|   |  | _      |                            |  |                             |
| OWIN Links  |  |        | Nucleotide Links           |  |                             |

|   |  |   | 01112-020   |             |                   |                             |
|---|--|---|-------------|-------------|-------------------|-----------------------------|
| S NCBI  | ENTREZ   | Nucl cotigle Poly morphism                        | JE T        |             | Welcome m         | My NCE<br>rrieder, (Sign Ou |
| Al Databases  | PubMed Nucleotide  | Protein Ge  | nome Shuct  | ture Poper  | at Taxonomy       | SNP                         |
| Search SNP  | : for  |   | 60 (Cea     |             |                   |                             |
|   | Limits Preview/Index Hist  | ory Clipboard Deta                                | is          |             |                   |                             |
| Entrez SNP<br>Search SNP<br>Search Mouse SNP<br>Common Query Filters                | or use search field tags en<br>• Boolean operators AND,<br>Function class: |   |             | clear       | Has genotype:     | clear                       |
| Entrez Batch Query<br>SNP Link Datamodel  | coding nonsynonymous<br>coding synonymous                                  | reference   | exception   | intron      | ☐ false<br>✓ true |                             |
| My NCBI<br>My NCBI help   | Chromosome(s):   |   |             | clear       | Map weight:       | clear                       |
|   | 919293 9   |   | □ 7         | □ 8<br>□ 16 |                   |                             |
| Entrez SNP Help<br>Search Fields<br>Programming Utilities<br>Batch Report<br>Legend | 9 9 10 9 11 9 1  | 2 □ 13 □ 14<br>0 □ 21 □ 22<br>Base Position: from | 0 15<br>0 W | ⊟ 10<br>⊟ X | 3-10<br>10+       |                             |



| S NC                         | 31 ENTREZ SING OF THE PART IN CONTRACT OF THE PART INCLUDE OF THE PART INC | My N<br>Welcome mnieder. Son  |
|------------------------------|--|---|
| Al Dorbertos<br>Search ( SNP | :) for   | Genome Shuchare Popol Taxonomy SNP<br>Go Clear  |
|                              | Search using wild-card(*), ranging(:), AND, OR, and NOT operat   |   |
|                              | Example<br>BRC*[Gene Name]   | Description<br>Search SNPs on all genes with names starting with the letter 'BRC' (ie. BRCA1 and<br>BRCA2)  |
|                              | [1:5[HET]<br>coding nonsynon[FUNC] AND 1[CHR]  | Search SNPs with heterozygosity between 1 and 5 percent<br>Search SNPs with function class 'coding nonsynon' located on chromosome 1  |
|                              | IICHRI OR 2[CHR]<br>IICHRI OR 2[CHR] NOT unknown[METHOD]<br>IIWEIGHTI AND (IICHRI OR 2[CHR]) NOT (unknown[METHOD   | Search all SNPs on chromosome 1 or 2<br>Search all SNPs on chromosome 1 or 2 detected by all methods except 'unknown'.<br>20 OR Search all SNPs with weight 1 on chromosome 1 or 2 detected by all methods except |
|                              | computed[METHOD])  | 'unknown' or 'computed'.  |
|                              |  |   |
|                              |  |   |

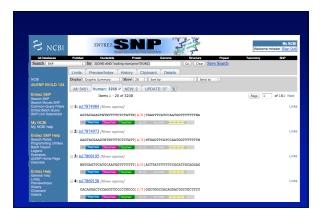
| Search Field           | Qualifier                                | Туре     | Description  |   |
|------------------------|--|----------|--|---|
| Allele                 | [ALLELE],[VARIATION],<br>[VARI]          | IUPAC    | Observed allele(s)<br>Example: N(ALLELE)   |   |
| Chromosome             | [CHR]                                    | Textnum  | Mapped chromosome number<br>Available values [1-22,W-Z, and<br>Example: 2[CHR] or X[CHR] | i Un (unknown)]   |
| Base Position          | [CHRPOS],[BPOS]                          | Integer* | Mapped chromosome position; u<br>field [CHR]<br>Example: 7[CHR] AND 885563               | use in conjunction with chromosome 398:88580839[CHRPOS] |
| Create Build<br>ID     | [CREATE_BUILD].[CBID]                    | Integer* | SNP create build ID<br>Example: 103[CBID]  |   |
| Publication<br>Date    | [CREATEDATE].[CDAT].[PDAT].<br>[PUBDATE] | Date*    | SNP create/publication date<br>Use the format YYYY/MM/DD<br>Example: 2002/01[CDATE]      | ; month and day are optional.                           |
| Function Class         | [FXN_CLASS], [FUNC]                      | Text     | Function Class:  |   |
|                        |  |          | locus region   | intron  |
|                        |  |          | coding nonsynon  | mma utr   |
|                        |  |          | coding synon<br>exception  | reference<br>splice site                                |
|                        |  |          | Example: "coding synon" [FUN4  | 9   |
| Gene Name              | [GENE], [GENE_SYMBOL]                    | Textnum  | Locus link symbol<br>Example: LPL[GENE]  |   |
| Protein<br>Description | [PDSC]                                   | Text     | Protein description<br>Example: "liver expressed" [PDS                                   | 6C]   |
| Gene<br>Description    | [GDSC]                                   | Text     | Gene description<br>Example: "kinase"[Gene Description]                                  | ption]  |
| GO Terms               | [GO]                                     | Text     | Gene Ontology (GO) terms<br>Example: "receptor"[GO term]                                 |   |
| Genotype               | [GENOTYPE], [GTYPE]                      | Boolean  | Genotype<br>Example: true[GTYPE]   |   |

| Entrez SNP - Search Terms Fields |
|----------------------------------|
|----------------------------------|

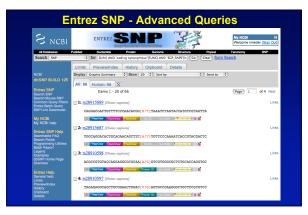
| Submitter<br>Handle  | [HANDLE]                   | Text     | Submitter handle<br>Example: "LEE"[HANDLE]  |
|----------------------|----------------------------|----------|---|
| Success Rate         | [SUCCESS_RATE], [SRATE]    | integer  | Success rate(%)<br>Example: <u>95[SRATE]</u>  |
| Organism             | [ORGN],[TAX_ID]            | Text     | Organism name or Taxonomy ID number; use the prefix 'txid' before<br>the Taxonomy ID number<br>Example: "human"[ORGN], ."mouse"[ORGN], or<br>'txid#566'IORGN] |
| Update Build<br>ID   | [UPD_BUILD],[UBID]         | Integer* | SNP update build ID<br>Example: 102[UBID]   |
| Modification<br>Date | [UDATE], [UDAT], [MODDATE] | Date*    | SNP modification/update date<br>Use the format YYYY/MM/DD; month and day are optional.<br>Example: 2002/01 [UDATE]  |
| Validation           | [VALIDATION]               | Text     | Validation status:<br>by cluster by submitter<br>by frequency no info   |
|                      |                            |          | Example: "by cluster" [VALIDATION]  |

More advanced queries:

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



| ubmitter<br>Iandle   | [HANDLE]                   | Text     | Submitter handle<br>Example: "LEE"[HANDLE]  |
|----------------------|----------------------------|----------|---|
| uccess Rate          | [SUCCESS_RATE], [SRATE]    | Integer* | Success rate(%)<br>Example: <u>95[SRATE]</u>  |
| Organism             | [ORGN],[TAX_ID]            | Text     | Organism name or Taxonomy ID number; use the prefix 'txid' before<br>the Taxonomy ID number<br>Example: "human' [ORGN], "mouse" [ORGN], or<br>"reide906" [ORGN] |
| Jpdate Build<br>D    | [UPD_BUILD],[UBID]         | Integer* | SNP update build ID<br>Example: 102[UBID]   |
| Modification<br>Date | [UDATE], [UDAT], [MODDATE] | Date*    | SNP modification/update date<br>Use the format YYYY/MM/DD; month and day are optional.<br>Example: 2002.01 [UDATE]  |
| Validation           | [VALIDATION]               | Text     | Validation status:<br>by cluster by submitter<br>by frequency no info   |
|                      |                            |          | Example: "by cluster"[VALIDATION]   |
| re adv               | vanced queries:            |          |   |



|                   | hai offi information for ge  | iotyping/onalaotorization              |
|-------------------|--|--|
| √<br>><br>A<br>[] | nat is the SNP? Flanking sequend<br>FASTA format<br>snp_name<br>CCGAGTAGCCAG<br>VG]<br>CTGGGATAGAAC  | ce and alleles.<br>EntrezSNP - better! |
| • dbs             | SNP reference SNP # (rs #) here is the SNP mapped? Exon, (   | promotor LITP, etc.                    |
| √<br>• Ho         | i picture of gene with mapped? Exon, i<br>picture of gene with mapped to t<br>w was it discovered? Method<br>hat assurances do you have that i | he gene structure.                     |
| • Wh              | nat population – African, European<br>nat is the allele frequency of each<br>e other SNPs associated - redund                                  | SNP? Common (>10%), rare               |

Minimal SNP information for genotyping/characterization

## Finding SNPs - Entrez SNP Summary

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

# Finding SNPs: Databases and Extraction

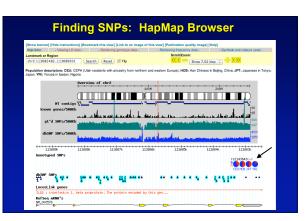
How do I find and download SNP data for analysis/genotyping?

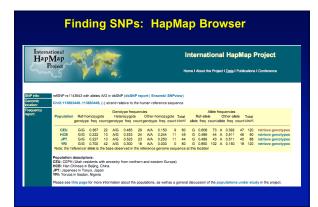
- 1. Entrez Gene - dbSNP
- Entrez SNP

# 2. HapMap Genome Browser

- 3. NIEHS Environmental Genome Project (EGP) Candidate gene website
- NIEHS web applications and other tools GeneSNPS, PolyDoms, TraFac, PolyPhen, ECR Browser, GVS

| International<br>HapMap<br>Project   | International HapMap Project   |
|--|--|
| Froject  | Home I About the Project I Data I Publications I Conference  |
| 中文   English   Français   日本語   Yorub  | a  |
| tates to develop a public resource that<br>iternational HapMap Project" for more     |  |
| Project Information  | News   |
| About the Project  | <ul> <li>2005-03-01: HapMap public release #16.</li> </ul>   |
| HapMap Publications  | ATTN: This the so-called Phase I data freeze which marks a major milestone of the project: a genotyped<br>common SNP every 5Kb in all populations under study. Data available for bulk download and graphical  |
| lapMap Conference<br>lapMap Mailing List   | browsing. Summary of genotyped SNPs:   |
| lapMap Mailing List<br>lapMap Project Participants                                   | Produktions CEU HCB JPT YBI  |
| lapMap Mirror Site in Japan  | Populations CEU HCB JPT TKI<br>Genotyped SNPs 1,073,863 1,044,888 1,044,416 1,034,205  |
| Project Data   | <ul> <li>2005-02-08: HapMap News Volume 1, 2004</li> <li>This is the first in a series of newsletters to be published by the Coriell Institute for Medical Research to</li> </ul>  |
| Seneric Genome Browser   | This is the first in a series of newsletters to be published by the Coriel Institute for Medical Hesearch to<br>inform communities how their samples are being used. Each issue of the newsletter will be available in the   |
| SUK Data Download  | primary languages of all the participating communities.  |
| NCODE Project  | <ul> <li>2005-02-07:International HapMap Consortium Expands Mapping Effort</li> </ul>  |
| Buidelines For Data Use  | 200 do of the memory and the provided 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 discover of onese related to common diseases, such as asthma, cancer. |
| Jseful Links   | The map will accelerate the discovery of genes related to common diseases, such as asthma, cancer,<br>diabetes and heart disease.  |
| lapMap Project Press Release   |  |
|  | Old News   |
|  |  |
| NHGRI HapMap Page<br>NCBI Variation Database (dbSNP)<br>Japanese SNP Database (JSNP) |  |





|  | Finding SNPs: HapMap Genotypes   |
|--|--|
| Fror details or<br>rs# allelss chr<br>Na10839 Na12875<br>Na12716 Na12807<br>Na10860 Na12044<br>Na10893 Na12264<br>Na12707 Na12056<br>Na12146 Na12762<br>rs1143643 A/G C<br>msgill-ggic.hag<br>AG AG AA AG AG | D1015 1085, Napage Tregency data days. T gras genetyped in population cru on chr2111161144,.11101444<br>Ella format, ene http://www.hepge.org/portpoge.<br>Aniiis sailis a hills a hills a hill be a start of the start<br>start of start of the star |
|  |  |
| Bulk data dow  | nloads   |
|  | tories contain project data that have been made publicly available. (See HapMap Data Access Policy for more  |
|  | details about each dataset can be found in READMEs in the respective directories:  |
|  | details about each dataset can be found in HEADMEs in the respective directories:  |
| Genotype:     LD Data: L     Allocated     criteria (se     Frequenci     been subm  |  |
| Genotype:     LD Data: L     Allocated     criteria (se     Frequenci     been subm     SNP assay     genotyping     Protocols   | Individual genotype data submitted to the DCC to date.<br>Indege disequilibrium properties D', LOD, P <sup>2</sup> compiled from the genotype data to date<br>DRs: dcbNP reference SNP clusters that have been priced and prioritized for genotyping according to several<br>info on how SNPa were selected). The file ORFEADME contains per-chromosome SNP counts and further details.<br>s: Allek S genotype frequencies compiled from genotyping data submitted to the DCC to date. These have also<br>tited to dbSNP and should be available in the next dbSNP build.  |

| Genotyped SNP:  | 113688k   | 113687k 11  | 3696k 11.3695k  | 113684k 113683k             |
|---|---|---|---|-----------------------------|
|   |   |   |   | CEUHCE JPT VRI              |
| LocusLink gen   | * *****   | · · · · · · · · · · · · · · · · · · ·   | ** * * * * *  |                             |
|   |   | in. The protein encoded by  | this sen  | ······ ·                    |
| RefSeq mRNA's   |   |   |   |                             |
| NM_000576   |   | of cenctype data install Hapky  | Annotate LD Plot<br>Dump Decorated FASTA File<br>Dump HapMap GFF File   |                             |
|   |   |   |   |                             |
|   |   |   | Dump HapMap LD Data   |                             |
| laploview (ver3.0) is nov<br>Data Source  | v available for download.   |   | Dump HapMap LD Data<br>Dump SNP frequency data  | 16:                         |
| taploview (ver3.0) is nov<br>Data Source<br>TapMap Data Rel#16/phar   | v available for download.   |   | Dump HapMap LD Data<br>Dump SNP frequency data  | hs:<br>About) Configure) Co |
| taploview (ver3.0) is nov<br>Data Source<br>HapMap Data Rel#16/phar<br>Tracks [Hide]                              | v available for download.   |   | Dump HapMap LD Data<br>Dump SNP frequency data  |                             |
| taploview (ver3.0) is nov<br>Data Source<br>HapMap Data Rel#16/pha  | v available for download.<br>sel Mar05, on NCBI B34   | assembly, dbSNP b122  | Dump HapMap LD Data<br>Dump SNP frequency data  |                             |
| lapioview (ver3.0) is nov<br>Data Source<br>IapMap Data Re#16/phar<br>Tracks [Hide]<br>External tracks italicized | vavailable for download.<br>sel Mar05, on NCBI B34 :<br>Contigs                                 | assembly, dbSNP b122  | Dump HapMap CPF Hite<br>Dump HapMap LD Data<br>Dump SNP frequency data  |                             |
| lapioview (ver3.0) is nov<br>Data Source<br>IapMap Data Re#16/phar<br>Tracks [Hide]<br>External tracks italicized | v available for download.<br>sel Mar05, on NCBI B34 :<br>Contigs<br>CYT:overview*               | essembly, dbSNP b122<br>Genotyped SNPs<br>gt/d SNPs/500Kb1<br>Heterozygosity/500Kb1 | Dump HapMap LD Data<br>Dump SNP frequency data<br>Ump SNP genotype data<br>Highlight SNP Properties   |                             |
| taploview (ver3.0) is nov<br>Data Source<br>IapMap Data Re#16/phar<br>Tracks [Hide]                               | r available for download.<br>sel Mar05, on NCBI B34 :<br>Contigs<br>CYT:overview*<br>dbSNP SNPs | essembly, dbSNP b122<br>Genotyped SNPs<br>gt/d SNPs/500Kb1<br>Heterozygosity/500Kb1 | Dump HapMap LD Data     Dump SNP frequency data     Dump SNP genotype data     Highlight SNP Properties     progenicus Prot     RefSeq mRNA's     Sequence Tagged Sites |                             |

Finding SNPs: HapMap Browser

| <ul> <li>What is the SNP</li> <li>✓ FASTA form</li> </ul> | ? Flanking sequence and alleles.              |
|---|---|
| >snp_name   |   |
| ACCGAGTAG   | CCAG  |
| [A/G]<br>ACTGGGATAG                                       | GAAC  |
| dbSNP reference   | e SNP # (rs #)                                |
| • Where is the SN   | P mapped? Exon, promoter, UTR, etc            |
|   | ne with mapped to the gene structure.         |
| How was it disco  |   |
|   | s do you have that it is real? Validated how? |
|   | – 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

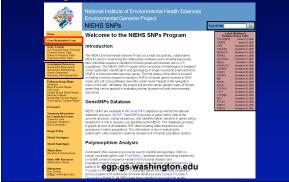
- 1. HapMap data sets are useful because individual genotype data can be used to determine optimal genotyping strategies (tagSNPs) or perform population genetic analyses (linkage disequilbrium)
- 2. Data are specific produced by those projects (not all dbSNP)
  - ✓ HapMap data is available in dbSNP
- 3. HapMap data (Phase II) can be accessed preleased 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?

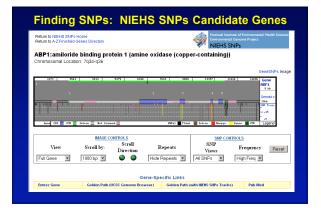
- 1. Entrez Gene - dbSNP - Entrez SNP
- 2. HapMap Genome Browser
- 3. NIEHS Environmental Genome Project (EGP) Candidate gene website
- 4. NIEHS web applications and other tools GeneSNPS, PolyDoms, TraFac, PolyPhen, ECR Browser, GVS

# Finding SNPs: NIEHS SNPs Candidate Genes



## Finding SNPs: NIEHS SNPs Candidate Genes

|  |                   | stitute of Environmental Health Sciences<br>ntal Genome Project<br>NPs                 | Swarch Site | 60            |
|--|-------------------|--|-------------|---------------|
| Reno   | Finished          | Genes  |             |               |
| Gene Nomination Form   | Hupo Gene<br>Nama | Gene Description   | Entrez Gene | Gene<br>Panel |
| Pintshed Genes Table<br>Genes in Progress Table<br>Program Genes Table                         | ADCB1             | ATP-binding cassette, sub-family B (MCR/TAP), member 1                                 | 5243        | <u>82</u>     |
| Genotyping Researces<br>Genotyping Decksround  | <u>A0.1</u>       | V-abl Abelson murine leukemia vital oncogene homolog 1                                 | 25          | <u>P2</u>     |
| Genotyping Dackground<br>Genotyping Methodalogy<br>Genes Genotyping<br>Sammary Genetyping Data | A0.2              | V-abl Abelson murine leukernia vital oncogene homolog 2 (arg.<br>Abelson-related gene) | 22          | E2            |
| Pathway Image Maps<br>Appendix<br>Pate Excision Recail   | ABP1              | Amilori de binding protein 1 (amine cuidase (copper-containing))                       | 25          | 82            |
| Call Cycle<br>Call Cycle<br>Double Strand Break Repair<br>Microsofth Danair                    | ACTE              | Actin, beta  | 92          | <u>P1</u>     |
| Misristich Repair<br>Nucleatide Existics Repair<br>Transcription Coupled Repair                | ACHIA             | Alcohol dehydrogenase 1A (class I), alpha polypeptide                                  | 124         | <u>P2</u>     |
| Personal   | ACH18             | Alcohol dehydrogenase IB (class I), beta polypept de                                   | 125         | <u>P2</u>     |
| Summary Information<br>for Completed Genes   | ADHIC             | Alcohol dehydrogenase 1C (class I), gamma polypeptide                                  | 126         | <u>P2</u>     |
| Semmary Data<br>Semmary Datatics<br>Data Downloads   | ACHH              | Alcohol dehydrogenese 4 (class II), pi połypeptide                                     | 127         | <u>P2</u>     |
| Usage Policy   | ACH6              | Alcohol dehydrogenese 5 (class II), chi polypeptide                                    | 128         | <u>P2</u>     |
| Viseal Geostypes   | ACH6              | Alcohol dehydrogenese 6 (class V)  | 150         | <u>P2</u>     |
| Visual Replotypes  | ADM               | Adrenomedulin  | 123         | <u>P2</u>     |
| What's New<br>Recently Finished Genes  | ADPRT             | ADP-ribosyltransferase (NAD+; poly (ADP-ribose) polymerase)                            | 142         |               |
| Other SBP Resources<br>Information Genes<br>Webmenter  | ACPRTL2           | ADP-ribosyltransferase (NAD+; poly(ADP-ribose) polymerase)-like 2                      | 10038       | <u>P1</u>     |
| Webmaster<br>Elio Tostev<br>Gritting, washington, edul   | AHR               | Anyl hydrocarbon receptor  | 196         | <u>P1</u>     |



# Finding SNPs: NIEHS SNPs Candidate Genes

| gattaac<br>ttattgca<br>gattaac | Mapping Data           | cSNPs<br>cDNA                              | Color FASTA<br>SNP Context          | PCR Primers (FASTA)<br>Genbank |
|--------------------------------|------------------------|--|-------------------------------------|--------------------------------|
| <u>.</u>                       | Genotyping Data        | Visual Genotype<br>Individual Genotypes    | SNP Alleles<br>SNP Allele Frequency | SNP Hardy-Weinberg             |
| e.                             | Haplotyping Data       | PHASE Output<br>Visual Haplotype           | Phased Individual<br>Haplotypes     | Sorted by Frequency            |
| $L_D$                          | Linkage Data           | LD Select (Tag SNPs)<br>African Descent Eu | iropean Descent Hispan              | ic Descent Asian Descent       |
| ??                             | Predictive<br>Analyses | Nonsynonymous cSNP Ar                      | nalysis                             |                                |

| gattaa            | 1        |             | [        | cSNPs   |         |            |           | olor FA |         |            | PCR P        |        | (FAST  | A)     |         |
|-------------------|----------|-------------|----------|---------|---------|------------|-----------|---------|---------|------------|--------------|--------|--------|--------|---------|
| ttattgo<br>gattaa | Mapp     | bing Data   |          | CDNA    |         |            | SI        | NP Con  | text    |            | Genba        | nk     |        |        |         |
| Site              | CON4-Pos | Description | Allelel  | AD-Free | ED-free | Total-free | AlleleZ   | AD-Free | ED-free | Total-free | Protein-Pos  | Ceden1 | Codoo7 | Aninol | laino2  |
| 989392            | 95959    | Description | Atterer  | 9.85    | 0.86    | 9.87       | 6 ALLELEZ | 8.12    | 0.14    | 8.13       | 8989         | Coooni | Cooone | Anthol | America |
| 888588            | 82828    |             | ÷.       | 8.91    | 1.99    | 8,95       |           | 8.89    | 0.00    | 8.84       | 2020         |        |        |        |         |
| 989794            | 82828    |             | ć.       | 8 48    | 8 82    | 8.68       | ÷.        | 9.68    | 8 18    | 8 48       | 5959         |        |        |        |         |
| 223959            | 82929    |             | ÷.       | 8.98    | 1.99    | 8.99       | é.        | 8.82    | 0.00    | 8.81       | 0000         |        |        |        |         |
| 981274            | 92929    |             | Ť        | 8.38    | 0.83    | 8.68       | ĉ         | 8.62    | 8.17    | 8.48       | 8989         |        |        |        |         |
| 981455            | 92929    |             |          | 8.98    | 1.00    | 0.99       |           | 9.92    | 0.00    | 0.01       | 0000         |        |        |        |         |
| 981913            | 95959    |             | c        | 8.92    | 1.00    | 8.95       | т         | 9.98    | 0.00    | 8,84       | 6989         |        |        |        |         |
| 981942            | 92929    |             | ¢ .      | 8.98    | 1.00    | 0.99       | 6         | 9.92    | 0.00    | 0.01       | 0000         |        |        |        |         |
| 881964            | 95959    |             | Т        | 8.98    | 1.00    | 8.99       | c         | 8.82    | 0.00    | 0.01       | 6666         |        |        |        |         |
| 982878            | 92929    |             | A        | 8.98    | 1.00    | 0.99       | G         | 8.82    | 0.00    | 0.01       | 6959         |        |        |        |         |
| 082143            | 95959    |             | Т        | 8.56    | 0.82    | 0.68       | c         | 8.44    | 0.18    | 0.32       | 6666         |        |        |        |         |
| 982791            | 95959    |             | 6        | 8.96    | 1.00    | 0.98       | т         | 0.04    | 0.00    | 0.02       | 6959         |        |        |        |         |
| 982712            | 95959    |             | A        | 8.91    | 1.00    | 8.95       | т         | 8.89    | 8.88    | 8.84       | 6666         |        |        |        |         |
| 982714            | 92929    |             | 6        | 8.98    | 1.00    | 0.99       | Α         | 0.02    | 0.00    | 0.01       | 6959         |        |        |        |         |
| 882766            | 95959    |             |          | 8.84    | 0.24    | 8.14       | +         | 8.96    | 0.76    | 0.85       | 8989         |        |        |        |         |
| 982824            | 95959    |             | т        | 8.98    | 1.00    | 8.99       | c         | 0.02    | 0.00    | 8.81       | 6989         |        |        |        |         |
| 003298            | 95959    |             | 6        | 8.69    | 0.43    | 8.56       | A         | 0.31    | 0.57    | 8.44       | 6959         |        |        |        |         |
| 883535            | 95959    |             | т        | 1.08    | 0.98    | 8.99       | A         | 8.88    | 0.02    | 8.81       | 8888         |        |        |        |         |
| 994996            | 95959    |             | <u>c</u> | 8.73    | 0.98    | 0.85       | <u> </u>  | 8.27    | 0.02    | 0.15       | 9959         |        |        |        |         |
| 984868            | 95959    |             | т        | 8.96    | 1.00    | 8.98       | c         | 8.84    | 0.00    | 8.82       | 8888         |        |        |        |         |
| 894978            | 92929    |             |          | 8.97    | 1.00    | 8.99       | +         | 8.83    | 0.00    | 0.01       | 6959         |        |        |        |         |
| 984246            |          |             |          | 8.98    | 1.00    | 8.99       |           | 8.82    | 0.00    | 8.81       | 8888         |        |        |        |         |
| 884325            | 95959    |             | c        | 8.98    | 0.78    | 0.88       | 1         | 8.82    | 0.22    | 0.12       | 6959         |        |        |        |         |
| 884392<br>884585  | 92929    |             | 1        | 8.88    | 0.78    | 8.79       | 6         | 8.12    | 0.30    | 0.21       | 5959<br>5959 |        |        |        |         |
| 985298            | 92929    |             | 2        | 8.98    | 0.55    | 0.99       |           | 8.82    | 0.00    | 8.81       | 2020         |        |        |        |         |
| 985298            | 92929    | SYNON       | 2        | 8.88    | 0.55    | 8.65       | ÷         | 8.12    | 0.45    | 0.34       | 0105         | TTC    | TIT    | PHECED | PHECED  |
| 005277            | 82920    | STNON       | 2        | 8.88    | 1.00    | 0.82       | 4         | 8.96    | 0.25    | 0.18       | 0000         | THC .  |        | PHECED | MILCHY  |
| 886293            | 92929    |             | -        | 1.98    | 0.98    | 8.99       | 1         | 8.85    | 0.02    | 0.05       | 1929         |        |        |        |         |
| 006293            | 82828    |             | 6        | 9.84    | 0.98    | 8.99       | 2         | 8.00    | 0.26    | 0.01       | 2020         |        |        |        |         |
| 986873            | 92929    |             | 2        | 8.98    | 1.00    | 8.99       | ÷         | 8.82    | 0.20    | 0.01       | 2929         |        |        |        |         |
| 986874            | 22020    |             | 2        | 9.87    | 8.72    | 8.79       |           | 8.13    | 0.28    | 8.21       | 2020         |        |        |        |         |
| 9265995           | 92929    |             |          | 8.17    | 0.72    | 8.23       | 2         | 8.82    | 0.20    | 8.77       | 0000         |        |        |        |         |
| 887114            | 82828    |             | c .      | 8.75    | 0.96    | 0.85       | ÷.        | 8.25    | 0.04    | 8.15       | 6959         |        |        |        |         |
| 887135            | 93939    |             | č.       | 8.94    | 1.00    | 8.97       | ÷.        | 8.96    | 0.00    | 8.83       | 8989         |        |        |        |         |
| 887154            | 82828    |             | i i      | 1.98    | 8.98    | 8.99       | é.        | 8.88    | 8.82    | 8.81       | 8929         |        |        |        |         |
| 887275            | 92929    |             | č        | 1.98    | 0.98    | 8.99       | ě.        | 8.98    | 0.02    | 8.81       | 8989         |        |        |        |         |
| 887277            | 82828    |             | Á.       | 8.98    | 1.00    | 8,99       | ć.        | 8.82    | 0.00    | 8.81       | 8929         |        |        |        |         |
| 988234            | 81481    | 31 UTR      | 2        | 8.85    | 0.72    | 8,79       | ĉ         | 8.15    | 0.28    | 8.21       | 8989         |        |        |        |         |
| 888449            | 82828    |             | ÷        | 8.94    | 1.00    | 8.97       | é.        | 8.85    | 0.00    | 8.83       | 0000         |        |        |        |         |

| ttaac<br>ttgca<br>ttaac Mapping Dat                | cSNPs<br>a cDNA  | Color FASTA<br>SNP Context  | PCR Primers (FASTA)<br>Genbank   |
|--|--|---|--|
| Search Nucleocide                                  | Pathor Westerne<br>Pathor Westerne<br>Sof oy94896()<br>Limits Preventindes Histo |   | C) My AC(8 IV<br>C) (Sign in 2) Factoria<br>PMC Texnology Bolds<br>out Sive Search |
| About Entrez<br>Entrez Nucleotide<br>Help   FAQ    | Display Summery  | Show 20     Send to      RefSeq 0     K lucleotide (1), EST (0), GSS (0). | hat's this?1   |
| Entrez Tools<br>Check sequence<br>revision history | 1: <u>A7948960</u> Reports   | ing protein 1 (amine oxidase (copper-con                                  | Links  |
| Linküut<br>My NCBI                                 |  |   |  |

| gattaac             |                   | cS         | NPs        |                          | Coli       | or FA | STA                 | PCR Primers (FAST | A) |  |
|---------------------|-------------------|------------|------------|--------------------------|------------|-------|---------------------|-------------------|----|--|
| ttattgca<br>gattaac | tgca Mapping Data |            | cDNA       |                          | SNF        | , Cor | text                | Genbank           |    |  |
|                     |                   |            |            | GTTACAGOCT               |            |       |                     |                   |    |  |
|                     |                   |            |            | GOCCGAGGCG               |            |       | ALTER.              |                   |    |  |
|                     |                   |            |            | ACATOGTGAA               |            |       |                     |                   |    |  |
|                     |                   |            |            | GGCGGCAGGT               |            |       |                     |                   |    |  |
|                     |                   |            |            | TOOTOTGAAC               |            |       |                     |                   |    |  |
|                     |                   |            |            | GCACTCCAGC               |            |       |                     |                   |    |  |
|                     | CACTGAGACT        | CTGTCTCAAA | *****      | AAGTGTTATG               | ATGCAGACCT | 350   | var(302):[G:0.13]   |                   |    |  |
|                     |                   |            |            | CTCCAGGCAC               |            |       |                     |                   |    |  |
|                     | CTGGACTCTC        | ATTCATTCTA |            | GCTCCCCTGG               |            | 450   |                     |                   |    |  |
|                     | GGAGCAGGCA        |            |            | AAGAGAAACT               | GATAACTCTT | 500   | var(500):[:0.96]    |                   |    |  |
|                     |                   |            |            | ATCOATATTC               |            |       |                     |                   |    |  |
|                     |                   |            |            | ATCATTTAG                |            |       |                     |                   |    |  |
|                     |                   |            |            | TGGCTAGGGT               |            |       |                     |                   |    |  |
|                     |                   |            |            | ACCOTTOCCC               |            |       |                     |                   |    |  |
|                     |                   |            |            | TCTGCCCCAG               |            |       |                     |                   |    |  |
|                     |                   |            |            |                          |            |       | var(794):[7:0.40]   |                   |    |  |
|                     |                   |            |            | CCTCANTACA               |            |       |                     |                   |    |  |
|                     |                   |            |            |                          |            |       | var(855):[0:0.01]   |                   |    |  |
|                     |                   |            |            | GAAAAGTATG               |            |       |                     |                   |    |  |
|                     |                   |            |            | ATACTABATT               |            |       |                     |                   |    |  |
|                     |                   |            |            | TOGACATCAN               |            |       |                     |                   |    |  |
|                     |                   |            |            | GTGATACATT               |            |       |                     |                   |    |  |
|                     |                   |            |            | GAACCAGAGA               |            |       |                     |                   |    |  |
|                     |                   |            |            | ACCANTACTC<br>CAACAGAGAA |            |       |                     |                   |    |  |
|                     |                   |            |            | ACAGCGAGGG               |            |       |                     |                   |    |  |
|                     |                   |            |            | CAACCORDEE               |            |       |                     | u                 |    |  |
|                     |                   |            |            | GACTTTAATC               |            |       |                     |                   |    |  |
|                     |                   |            |            |                          |            |       |                     |                   |    |  |
|                     |                   |            |            |                          |            |       |                     |                   |    |  |
|                     |                   |            |            | AATGCCAGGT               |            |       | 1 401(1455)1(14155) |                   |    |  |
|                     |                   |            |            |                          |            |       |                     |                   |    |  |
|                     |                   |            |            |                          |            |       |                     |                   |    |  |
|                     |                   |            |            | AATGCTTTTC               |            |       |                     |                   |    |  |
|                     |                   |            |            | TACTCAGAGA               |            |       |                     |                   |    |  |
|                     |                   |            |            | AAATCACACA               |            |       |                     |                   |    |  |
|                     | COTCATOGOG        | AAGTCACTCA | TTTTCTCCTT | TTTACACAG                | TOTCTGAAGO | 1850  | Exon 2 UTR          |                   |    |  |
|                     |                   |            |            | TGAAATGATG               |            |       |                     |                   |    |  |
|                     |                   | E V P      | ELAS       | EMM                      | AYY        |       |                     |                   |    |  |

|                             | Visual Genotype<br>IndMidual Genotypes   | SNP Alleles<br>SNP Allele Frequency | SNP Hardy-Weinberg |
|-----------------------------|--|-------------------------------------|--------------------|
| African American            |  |                                     |                    |
| African YRI<br>European CEU |  |                                     |                    |
| Hispanic                    |  |                                     |                    |
| Asian CHB JPT               | 1993 - 18<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1<br>1  |                                     |                    |
|                             | VG2 of: abplx<br>memory are means at a fail<br>memory and a second at a<br>memory and at a<br>memory |                                     |                    |

| Genot | typing Data |         | Genotype<br>lual Genotypes | SNP Alleles<br>SNP Allele Frequency                   | SNP Hardy-Weinberg |
|-------|-------------|---------|----------------------------|---|--------------------|
|       | 000302      |         | λ                          |   |                    |
|       | 000302      |         | A A                        |   |                    |
|       | 000302      |         | <b>л</b>                   |   |                    |
|       | 000302      |         | A A                        |   |                    |
|       | 000302      |         | λ λ<br>λ                   |   |                    |
|       | 000302      |         | ι λ<br>                    |   |                    |
|       | 000302      |         | ι <i>Α</i>                 |   |                    |
|       | 000302      |         | λ λ                        |   |                    |
|       | 000302      |         | λ λ                        |   |                    |
|       | 000302      | D011 J  | ۱ C                        |   |                    |
|       | 000302      |         | A A                        |   |                    |
|       | 000302      |         | A G                        |   |                    |
|       | 000302      |         | ι <u>λ</u>                 |   |                    |
|       | 000302      |         | G                          |   |                    |
|       | 000302      |         | A A G                      |   |                    |
|       | 000302      |         | 1 6                        |   |                    |
|       | 000302      |         | ι ο<br>\ λ                 |   |                    |
|       | 000302      |         |                            |   |                    |
|       | 000302      |         | λ                          |   |                    |
|       | 000302      | D038 2  | ۱ C                        |   |                    |
|       | 000302      |         | A A                        |   |                    |
|       | 000302      | D040 2  | A A                        |   |                    |
|       |             |         |                            | tab> allele1 <ta< th=""><th>ab&gt; allele2</th></ta<> | ab> allele2        |
| Rep   | peat for    | r all i | ndividual                  | S   |                    |
|       |             |         |                            |   |                    |
| Rep   | peat for    | r nex   | I SNP                      |   |                    |

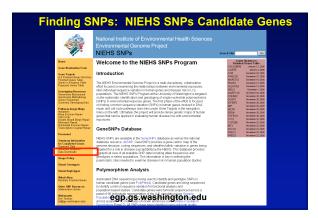
|        |         | Genotypin | <u>g Data</u> |        | Genotyp<br>lual Geno | -         |         | Alleles<br>Allele Fre | quency | <u>SNP</u> | Hardy-We | <u>einberg</u> |       |
|--------|---------|-----------|---------------|--------|----------------------|-----------|---------|-----------------------|--------|------------|----------|----------------|-------|
| Site   | Allelei | AD-pop    | go-pop        | HD-pop | XD-pop               | Total-pop | Allele2 | AD-pop                | E3-pop | HD-pop     | ID-pop   | Total-pop      | AD-hz |
| 000065 | A       | 0.02      | 0.00          | 0.00   | 0.00                 | 0.01      | G       | 0.90                  | 1.00   | 1.00       | 1.00     | 0.99           | 0.04  |
| 000089 |         | 0.00      | 0.02          | 0.00   | 0.00                 | 0.01      | ġ.      | 1.00                  | 0.98   | 1.00       | 1.00     | 0.99           | 0.00  |
| 000111 | A       | 0,00      | 0,02          | 0.00   | 0.00                 | 0.01      | 0       | 1.00                  | 0,98   | 1.00       | 1,00     | 0,99           | 0,00  |
| 000310 | 6       | 0.52      | 0.27          | 0.32   | 0.04                 | 0.39      | ٨       | 0.48                  | 0.73   | 0.68       | 0.96     | 0.71           | 0.50  |
| 000537 | -       | 0.00      | 0.00          | 0.00   | 0.04                 | 0.01      | +       | 1.00                  | 1.00   | 1.00       | 0.96     | 0.99           | 0.00  |
| 000614 | с       | 0.17      | 0.00          | 0.00   | 0.00                 | 0.05      | т       | 0.83                  | 1.00   | 1.00       | 1.00     | 0.95           | 0.25  |
| 000677 | *       | 0.02      | 0.07          | 0.00   | 0.00                 | 0.02      | 0       | 0.98                  | 0.93   | 1.00       | 1.00     | 0.98           | 0.04  |
| 000775 | с       | 0.52      | 0.27          | 0.32   | 0.04                 | 0.29      | т       | 0.48                  | 0.73   | 0.68       | 0.96     | 0.71           | 0.50  |
| 000873 | *       | 0.02      | 0.00          | 0.00   | 0.00                 | 0.01      | G       | 0.98                  | 1.00   | 1.00       | 1.00     | 0.99           | 0.04  |
| 000940 | *       | 0.02      | 0.16          | 0.15   | 0.04                 | 0.10      | G       | 0.98                  | 0.84   | 0.82       | 0.96     | 0.90           | 0.04  |
| 001314 | T       | 0.26      | 0.26          | 0.32   | 0.04                 | 0.22      | G       | 0.74                  | 0.74   | 0.68       | 0.96     | 0.78           | 0.38  |
| 001574 | -       | 0.04      | 0.00          | 0.00   | 0.00                 | 0.01      | +       | 0.96                  | 1.00   | 1.00       | 1.00     | 0.99           | 0.07  |
| 001902 | Ŧ       | 0.00      | 0.00          | 0.00   | 0.06                 | 0.03      | c       | 1.00                  | 1.00   | 1.00       | 0.99     | 0.98           | 0.00  |
| 002159 | *       | 0.04      | 0.00          | 0.00   | 0.00                 | 0.01      | G       | 0.96                  | 1.00   | 1.00       | 1.00     | 0.99           | 0.07  |
| 002660 | *       | 0.19      | 0.25          | 0.32   | 0.04                 | 0.19      | c       | 0.81                  | 0.75   | 0.68       | 0.96     | 0.81           | 0.30  |
| 002674 | *       | 0.24      | 0.25          | 0.32   | 0.01                 | 0.21      | c       | 0.76                  | 0.75   | 0.68       | 0.96     | 0.79           | 0.37  |
| 002734 | 6       | 0.24      | 0.27          | 0.32   | 0.04                 | 0.22      | ٨       | 0.76                  | 0.73   | 0.68       | 0.96     | 0.78           | 0.37  |
| 002751 | G       | 0.07      | 0.00          | 0.00   | 0.00                 | 0.02      | *       | 0.93                  | 1.00   | 1.00       | 1.00     | 0.98           | 0.14  |
| 002776 | *       | 0.26      | 0.00          | 0.00   | 0.00                 | 0.07      | т       | 0.74                  | 1.00   | 1.00       | 1.00     | 0.93           | 0.35  |
| 003800 | *       | 0.07      | 0.00          | 0.00   | 0.00                 | 0.02      | a .     | 0.93                  | 1.00   | 1.00       | 1.00     | 0.98           | 0.14  |
| 004015 | A.      | 0.06      | 0.00          | 0.00   | 0.00                 | 0.02      | 0       | 0.94                  | 1.00   | 1.00       | 1.00     | 0.98           | 0.10  |
| 004266 | *       | 0.00      | 0.02          | 0.00   | 0.00                 | 0.01      | G       | 1.00                  | 0.98   | 1.00       | 1.00     | 0.99           | 0.00  |
| 004469 | -       | 0.57      | 0.23          | 0.29   | 0.05                 | 0.25      | +       | 0.42                  | 0.78   | 0.71       | 0.95     | 0.72           | 0.49  |
| 004837 | G       | 0.26      | 0.00          | 0.00   | 0.00                 | 0.07      | *       | 0.74                  | 1.00   | 1.00       | 1.00     | 0.93           | 0.38  |
| 004854 | *       | 0.02      | 0.00          | 0.00   | 0.00                 | 0.01      | 0       | 0.98                  | 1.00   | 1.00       | 1.00     | 0.99           | 0.01  |
| 005006 | 9       | 0.02      | 0.00          | 0.00   | 0.00                 | 0.01      | ٨       | 0.98                  | 1.00   | 1.00       | 1.00     | 0.99           | 0.04  |
| 005156 | T       | 0.60      | 0.30          | 0.30   | 0.04                 | 0.32      | с       | 0.40                  | 0.70   | 0.70       | 0.96     | 0.68           | 0.40  |
| 005215 | Ģ       | 0.05      | 0.00          | 0.00   | 0.00                 | 0.02      | 8       | 0.92                  | 1.00   | 1.00       | 1.00     | 0.95           | 0.15  |
| 005307 | T       | 0.68      | 0.29          | 0.22   | 0.01                 | 0.28      | *       | 0.32                  | 0.71   | 0.78       | 0.96     | 0.72           | 0.44  |
| 005429 | с       | 0.04      | 0.00          | 0.00   | 0.00                 | 0.01      | т       | 0.96                  | 1.00   | 1.00       | 1.00     | 0.99           | 0.08  |
| 005472 | T       | 0.00      | 0.00          | 0.02   | 0.00                 | 0.01      | с       | 1.00                  | 1.00   | 0.98       | 1.00     | 0.99           | 0.00  |
| 005720 | т       | 0.39      | 0.25          | 0.27   | 0.04                 | 0.24      | G       | 0.61                  | 0.75   | 0.73       | 0.96     | 0.76           | 0.45  |
| 005811 | T       | 0.04      | 0.00          | 0.00   | 0.00                 | 0.01      | c       | 0.96                  | 1.00   | 1.00       | 1.00     | 0.99           | 0.07  |
| 005817 | T       | 0.13      | 0.00          | 0.00   | 0.00                 | 0.01      | ¢       | 0.87                  | 1.00   | 1.00       | 1.00     | 0.96           | 0.23  |
| 005947 | 7       | 0.61      | 0.30          | 0.26   | 0.04                 | 0.31      | ¢       | 0.39                  | 0.70   | 0.74       | 0.96     | 0.69           | 0.48  |
| 006794 | T       | 0.09      | 0.00          | 0.00   | 0.00                 | 0.03      | c       | 0.91                  | 1.00   | 1.00       | 1.00     | 0.97           | 0.17  |
| 004895 | т       | 0.09      | 0.07          | 0.11   | 0.00                 | 0.07      | c       | 0.91                  | 0.93   | 0.89       | 1.00     | 0.93           | 0.17  |
| 006926 | T       | 0.02      | 0.05          | 0.05   | 0.01                 | 0.01      | c       | 0.98                  | 0.95   | 0.95       | 0.96     | 0.96           | 0.01  |

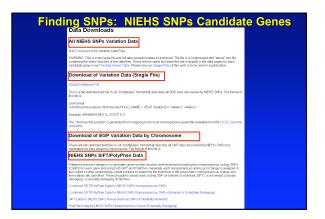
| Haplotyping Date           | a PHASE Output      | Phased Individual<br>Haplotypes | Sorted by Frequency |
|----------------------------|---------------------|---------------------------------|---------------------|
|                            | Visual Haplotype    |                                 |                     |
| EGIN LIST SUMMARY          |                     | SECIN BESTPAIRS_SUMMARY         |                     |
| 1 ADTCCCAIGACTGCCAGDCCCTT  | CTCGGGACCC 1.000000 | 0001: (1,4)                     |                     |
| 2 ADTCCCAIGACTGCCGGITCGT   | 7606666666 3.000000 | 0002: (13,22)                   |                     |
| 3 ADTCCCAIGACTGCCGGITCOT   | TOTOGGCOGC 2.000000 | 0003: (2,4)                     |                     |
| 4 ADTCCCAIGACTGTCAADCCCTC  |                     | D0061 (10,17)<br>D005: (11,22)  |                     |
| 5 ADTCCCAIGCCTGCCGGITCGT   | TCTCCCCCCC 1.000000 | 00061 (20.22)                   |                     |
| 6 ADTCCCAIGCCGGCCGGITCGT   |                     | 0007: (22.30)                   |                     |
| 7 ADTCCTAIGCCTGCTGGICCOT   |                     | 0008: (6,7)                     |                     |
| 8 ADTCCTAIGCCCGCCGGICCGTC  |                     | 0009+ (5,32)                    |                     |
| 9 ADTCCTTIGCCTGCCGGITCGTC  |                     | 0010: (4,13)                    |                     |
| 10 ADTTCCAIGACTGCCGGICCGCC | CGCGGACCCT 1.000000 | 0011: (23,34)                   |                     |
| 11 ADTICTAIGCCTGCCGGICCGT  |                     | 0012: (2,14) 0013: (7,33)       |                     |
| 12 ADTTCTAIACCTGTCGGICCGTC |                     | 0014: (8,22)                    |                     |
| 13 ADCCCTAIGCCTACCGGICCGTC |                     | 0015: (32,36)                   |                     |
| 14 ADCCCTTIGCCCGCCGGICCGTC | cccccccc 1.00000    | 0016: (10.31)                   |                     |
| 15 ADCTCTAIGCCTGCCGGICCGT  |                     | 00331 (7,34)                    |                     |
| 16 ADCTCTAIACCTGCCGGICCGTC |                     | 00341 (3,33)                    |                     |
| 17 ADCTCTAIACCTGCCGGICCGTC |                     | 0035: (4,28)                    |                     |
| 18 ADCTCTAIACCTGCCGGICCGCC |                     | 00361 (3,22)                    |                     |
| 19 ADCTCTAIACCTGCCGGITCGT  |                     | 0037: (9,9)                     |                     |
| 20 ADCTCTAIACCTGTCAADCCCTG |                     | 0038: (24,33)<br>0039: (9,32)   |                     |
| 21 ADCTCTAIACCTGTCGGICCGTC |                     | 00391 (9,32)<br>00401 (29,31)   |                     |
| 22 ADCTCTAIACCTACCGGICCGTC |                     | 2001: (22.22)                   |                     |
| 23 ADCTCTAIACCCGCCGGICCGT  |                     | 80021 (28,28)                   |                     |
| 24 ADCTCTAIACCCGCCGGICCGC  |                     | 2003: (22,33)                   |                     |
| 25 ADCICTAIACCCACCGGICCGTC |                     | 2004: (28,33)                   |                     |
| 26 ADCTCTADGCCCGTCGADCCCTC |                     | 8005+ (28,35)                   |                     |
| 27 ADCICTADGCTCGCCAADCCCTC |                     | 2006: (22,22)                   |                     |
| 28 ADCTCTADGCTCGTCAADCCCTC |                     | 2007: (22,28)<br>2008: (26,33)  |                     |
| 29 ADCTTTAIACCTGCCGGICCGT  |                     | 20081 (26,33)                   |                     |
| 30 ADCTTTAIACCTACCGGICCGTC |                     | 2010: (22,22)                   |                     |
| 31 ADCTTTAIACCTACCGGICCGTC |                     | 8011+ (19,22)                   |                     |
| 32 AITCCCAIGACTGCCGGITTGT  |                     | 0012: (22,33)                   |                     |
| 33 GDTCCCAIGCCTGCCGGICCGT  |                     | 2013: (16,22)                   |                     |
| 34 GDTCCCAIGCCTGCCGGICCGT  |                     | 2014; (2,22)                    |                     |
| 35 GDTCCCAIGCCCGTCAADCCCTT |                     | 2015: (22,25)                   |                     |
| 36 GDTCCCADGCCTGCCGGICCGTC | cccccccc 1.00000    | 8016: (15,22)<br>8017: (28,28)  |                     |

| Linkage Data   | LD Select (Tag SNPs)<br>African Descent European Descent Hispanic Descent Asian Descent  |
|--|--|
| Bin<br>Bin<br>Bin<br>Bin<br>Bin<br>Bin<br>Bin<br>Bin<br>Bin<br>Bin | <pre>Tradings: 1089<br/>Tradings: 1089<br/>Cotter_mps: 0189<br/>Cotter_mps: 0189<br/>Cotter_mps: 0189<br/>Cotter_mps: 0189<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:<br/>Cotter_mps:</pre> |

| Hugo<br>brcal | AllelePos<br>031631 | a.a.Pos<br>275 | Residue<br>G | Variant<br>S | Freq<br>0.01 | PPHpredict<br>benian | SIFTpredic:<br>TOLERATED |
|---------------|---------------------|----------------|--------------|--------------|--------------|----------------------|--------------------------|
| brca1         |                     | 356            | 0            | B            | 0.01         | possibly damaging    | INTOLERANT               |
| brcal         |                     | 693            | D            | ที่          | 0.05         | benion               | TOLERATED                |
| brcal         |                     | 871            | P            | L            | 0.43         | benign               | TOLERATED                |
| brcal         |                     | 1038           | E            | G            | 0.32         | benign               | INTOLERANT               |
| brcal         |                     | 1040           | S            | N            | 0.01         | benign               | TOLERATED                |
| brcal         |                     | 1140           | ŝ            | 6            | 0.02         | benign               | TOLERATED                |
| brcal         |                     | 1183           | ĸ            | R            | 0.33         | benign               | TOLERATED                |
| brcal         |                     | 1613           | s            | G            | 0.33         | benign               | TOLERATED                |
| brcal         |                     | 1620           | Ť            | Ā            | 0.01         | benion               | TOLERATED                |
| Γ = Soi       | ting Into           |                |              |              |              | nonymous SNP         | S                        |

| Download                       | a zip file of all data for th | s gene                                     |                                     | Sample Population Description  |
|--------------------------------|-------------------------------|--|-------------------------------------|--------------------------------|
| gattaac<br>ttattgca<br>gattaac | Mapping Data                  | CSNPs<br>CDNA                              | Color FASTA<br>SNP Context          | PCR Primers (FASTA)<br>Genbank |
|                                | Genotyping Data               | Visual Genotype<br>Individual Genotypes    | SNP Alleles<br>SNP Allele Frequency | SNP Hardy-Weinberg             |
|                                | Haplotyping Data              | PHASE Output<br>Visual Haplotype           | Phased Individual<br>Haplotypes     | Sorted by Frequency            |
| <sup>L</sup> <sub>D</sub>      | Linkage Data                  | LD Select (Tag SNPs)<br>African Descent Eu | ropean Descent Hispanic             | Descent Asian Descent          |
| 2                              | Predictive<br>Analyses        | Nonsynonymous cSNP Ar                      | nabesis                             |                                |





#### 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

| 0     | meSNI    | D 🕫            | My Gene ShiPy 💌 Game Lints | EOPLINE           | Toob              | - Open Query    |
|-------|----------|----------------|----------------------------|-------------------|-------------------|-----------------|
| Viere | tin Inc. | a Romane d 572 | MEHS<br>All Genes          |                   | ну <u>12/1</u> се | TANK RASTA TANK |
| 8     | - 10     | \$YMD01        | GENE MODION College        | LASS              | NUMBER            | LOCUS           |
| 8     | 5543     | ABCB1          | UCSO Hall Call Signal      | NOCESM            | 2                 | 2021.0          |
| 8     | 25       | AlILI          | UCSO MAL CNA Papel         | IMSON             | 2                 | 8434.3          |
| 2     | 27       | ABEA           | CONSIGNATION CONSIGNATION  | NIXINAL PRO       | 9                 | 1014-015        |
| 8     | 26       | A891           | UCSC by DET                | BRANDLEM          | 2                 | 2014-016        |
| 8     | 60       | ACTU           | UC80-ball6-1               | CELL_STRUCTURE    | 2                 | 7918-912        |
|       | 124      | ADHIA          | UCSCIMIC-                  | RETABOLISM        | 2                 | 4923-023        |
| 8     | 125      | ADHIP          | UC\$C \$416:4              | BETABOLISM        | 9                 | 49(1-9))        |
| 8     | 125      | ADHIC          | UC80 half-4                | BETADOLISM        | 2                 | 902-02          |
|       | 127      | ADR4           | UCSC halfed                | RETABOLISM        | 9                 | 9973-974        |
| 8     | 128      | ADHS           | UC\$C \$416.4              | BETABOLEM         | 2                 | 4921-925        |
| 8     | 130      | ADHS           | UC30.bal64                 | BETADOLEM         | 2                 | 942.2           |
| Ξ.    | 113      | ADM            | DOSCIMBER1                 | DOMEOSTASIS       | 2                 | 13915-4         |
| 8     | 142      | ADPRT          | UCSC 1416-1                | CRAS_BEINGLING    | 9                 | 1911-912        |
| 8     | 10035    | ADMIT12        | 11C8C-bg16-14              | DNA_RIPAR         | 2                 | 1911.1-011      |
| Ξ.    | 116      | ARR            | UCSC halfer                | HETAPOLISM        | 9                 | 7938            |
| 8     | 218      | AKT            | DCSC.3g16.19               | CREAL MERINALIPID | 2                 | 22410.1-111.2   |
| 8     | 200      | ALAD           | UC40.6416.9                | BREADOLDM         | 2                 | 2023            |
| 2     | 236      | ALDSUA1        | UC\$0.8416-9               | HETAPOLISM        | 2                 | 9421.33         |
| 2     | 8854     | ALDHIA2        | DOSC Male 15               | RETARCEDM         | 9                 | 15421-2         |
| 8     | 239      | ALOX12         | 10C8C-bg16:17              | BREASOLISM        | 2                 | 1711.1          |
| 2     | 354      | ADC2           | UCROMMETT                  | HETAPOLISM        | 2                 | 2740.2          |
| 2     | 8679     | A0C3           | DOSCIMBER7                 | CELL_MENALING     | 2                 | 17921           |
| 8     | 323      | APEXI          | 10080/hg16-14              | DISA_REPAR        | 2.0               | 1411.1-012      |
| 8     | 27391    | A19312         | UCSChalle-X                | DISA_REPAR        | 2                 | 2011.22         |

# GeneSNPs: One stop shopping



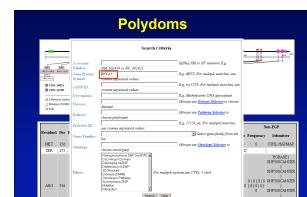


# Polydoms

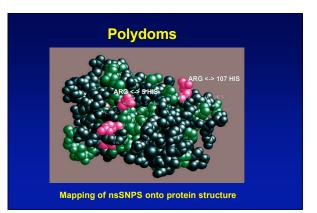
A web-based application that maps synonymous and non-synonymous SNPs onto known functional protein domains

- 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/



SNP5

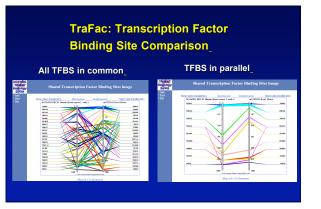


# TraFac: Transcription Factor Binding Site Comparison

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



## ECR Browser: Evolutionary Conserved Regions

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

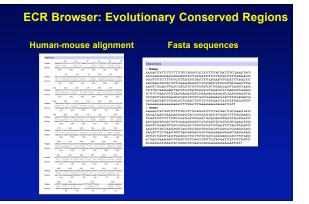
Gene annotations from UCSC Genome Browser

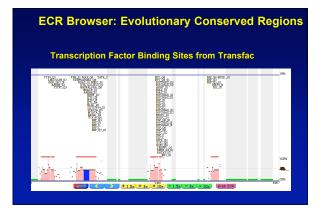
Easy retrieval of ECR sequences and alignments

Pre-computed transcription factor binding sites

http://ecrbrowser.dcode.org

#### ECR Browser: Evolutionary Conserved Regions ECR Browser on Human (hg17) Setting Mouse (n Rat (rn3) Frog (xt4) Fugu (fu3) Zebrafish 4.0 (zv4) Tetraodon (to6) Sisplay gene f mRNA Ensemble RefSeq Layer he Detect ECRs (Evolutionary C min length 100 bps min identity 70 % 30% 302% Relative C Abs \* Coor R select to show si Pugu (fu3) Chicken (gg2) Mouse (mm5) Mouse (mm6) Frog (xt3) 30% 5 30% m. 114 111





## PolyPhen: Polymorphism Phenotypingprediction of functional effect of human nsSNPs

- Physical and comparative analyses used to make predictions
- Uses SwissProt annotations to identify known domains
- Calculates a substitution probability from BLAST alignments of homologous and orthologous sequences
- Ranks substitutions on scale of predicted functional effects from "benign" to "probably damaging"

http://tux.embl-heidelberg.de/ramensky/

## PolyPhen: Polymorphism Phenotypingprediction of functional effect of human nsSNPs

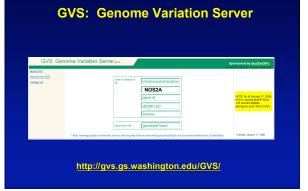


# **GVS: Genome Variation Server**

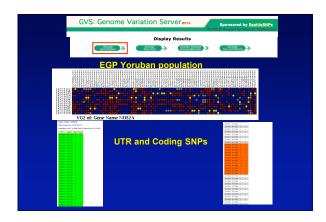
Provides rapid analysis of 4.3 million SNPs from dbSNP and the HapMap Mapped to human genome build 35 (hg17) Displays genotype data in text and image formats Displays tagSNPs or clusters of informative SNPs in text and image formats

Displays linkage disequilibrium (LD) in text and image formats

http://gvs.gs.washington.edu/GVS/

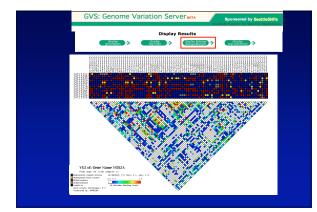


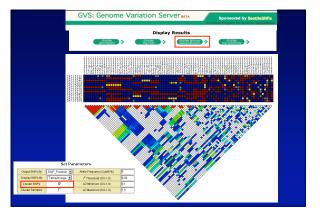
| Gene Name:    | NOS2A    |                      |                     |                                     |                         |               |                     |  |  |
|---------------|----------|----------------------|---------------------|-------------------------------------|-------------------------|---------------|---------------------|--|--|
| Gene ID: 4843 |          |                      | 1                   |                                     |                         |               |                     |  |  |
| Chromosome 1  | 7:231079 | 19 - 23151685        | 2(-) <u>Ele ma</u>  | ullene:                             |                         |               |                     |  |  |
|               |          |                      |                     |                                     |                         |               |                     |  |  |
|               |          | Charle of the        | stone constation W  | Select Po<br>u may select up to 4 p | opulation(s)            |               | instant of a second |  |  |
|               | Select   | Number<br>Variations | Number<br>Genotypes | Popula                              |                         |               | omitter             | Check to Select<br>Only Unrelated<br>Individuals |  |
|               | 17       | 225                  | 2800                | EOP_YORU                            | B-PANEL                 | por           | _SNPS               |  |  |
|               | E        | 235                  | 5515                | EOP_ASIA                            | HPANEL                  | DOP           | _SNPS               |  |  |
|               | Π.       | 236                  | 3518                | EOP_AD-                             | PHNEL                   | EOP_BNPS      |                     |  |  |
|               | E.       | 235                  | 5153                | EOP_CEPH                            | HPANEL                  | EGP           | SNP8                |  |  |
|               | E        | 283                  | 4428                | EGP_HISP                            | PANEL                   | EOP           | SNPS                |  |  |
|               | E        | 51                   | 4500                | Highlig                             | 2-1/70                  | <u>C5H</u>    | HAPMAP              | п  |  |
|               | Π.       | 50                   | 2250                | Hasting                             | SIPT                    | CEH           | HAPMAR              |  |  |
|               | E        | 50                   | 2250                | HapMap                              | HCB                     | CSH           | HAPMAP              |  |  |
|               | E.       | 48                   | 4320                | HapMap                              | -CEU                    | CSHL-HAPMAP   |                     | п  |  |
|               | E        | 21                   | 504                 | AID_CHN                             | PANEL                   | PERLEGEN      |                     |  |  |
|               | Π.       | 21                   | 483                 | ALD_AIR                             | PANE.                   |               |                     |  |  |
|               | E        | 21                   | 604                 | AFD_EUR                             | AFD_EUR_PWNEL           |               | LEGEN               |  |  |
|               |          |                      |                     | Set Pa                              | rameters                |               |                     |  |  |
|               |          |                      | Output SNPs By      | SNP_Position                        | Allele Frequen          | oy Catoff (%) | 1                   |  |  |
|               |          |                      | Display SNPs By     | Table/mage                          | r <sup>2</sup> Threshol | 8 (0.0-1.0)   | 0.80                |  |  |
|               |          |                      | Cluster SNPs        |                                     | LD Minimur              |               | 0.1                 |  |  |
|               |          |                      | Cluster Samples     | E                                   | LD Maximur              | n (0.0-1.0)   | 1.0                 |  |  |



|  | ency Cutof<br>HapMap-C<br>mple Aleie<br>20157 c<br>20158 c<br>20158 c<br>20158 c<br>20158 c<br>20158 c<br>20170 c<br>20171 c<br>20171 c<br>20171 c<br>20171 c   | t (%): )<br>EU, Sab  | mitter: CBH, HAPMAP | → display<br>tag snps | СЕРН |         |                                   |            |  |
|--|---|--|---------------------|-----------------------|------|---------|-----------------------------------|------------|--|
| Albih Firepart           Staff         Sar           2210845         140   | ency Cutof<br>HapMap-C<br>mple Aleie<br>20157 c<br>20158 c<br>20158 c<br>20158 c<br>20158 c<br>20158 c<br>20170 c<br>20171 c<br>20171 c<br>20171 c<br>20171 c   | f (%):1<br>EU, Sab<br>f  | mitter: CBH, HAPMAP | нармар                |      |         |                                   |            |  |
| Stat         Same           Stat         Same           2010005         140                              | HapNop C<br>mple Aleie<br>20157 c<br>20158 c<br>20158 c<br>20158 c<br>20158 c<br>20158 c<br>20171 c<br>20171 c<br>20171 c<br>20172 c<br>20178 c   | EU, Hab<br>1 Almo<br>6<br>6<br>6<br>6<br>6<br>6<br>6<br>6<br>6<br>6<br>6 | mitter: CBH, HAPMAP | нармар                |      |         |                                   |            |  |
| SHP         Sam           27100455         400           27100456         400 | Imple         Allele           00.153         0           00.153         0           00.155         0           00.156         0           00.156         0           00.156         0           00.156         0           00.156         0           00.156         0           00.156         0           00.156         0           00.156         0           00.156         0           00.171         0           00.105         0 | 2<br>4<br>6<br>6<br>6<br>6<br>6<br>6<br>6<br>6<br>6<br>7<br>6<br>7<br>6  |                     | нармар                |      |         |                                   |            |  |
| SHP         Sam           27100455         400           27100456         400 | Imple         Allele           00.153         0           00.153         0           00.155         0           00.156         0           00.156         0           00.156         0           00.156         0           00.156         0           00.156         0           00.156         0           00.156         0           00.156         0           00.156         0           00.171         0           00.105         0 | 2<br>4<br>6<br>6<br>6<br>6<br>6<br>6<br>6<br>6<br>6<br>7<br>6<br>7<br>6  |                     | нармар                |      |         |                                   |            |  |
| 2010655 140<br>2010655 140   | 201157 E<br>201158 E<br>201158 E<br>201158 E<br>201158 E<br>201158 E<br>201171 E<br>201171 E<br>201171 E<br>201172 E<br>201184 E  | 6<br>6<br>6<br>6<br>6<br>6<br>6  |                     |                       |      |         |                                   |            |  |
| 23382666 1427<br>23132655 1427   | 20150 E<br>20150 E<br>20160 E<br>20160 E<br>20160 E<br>20170 E<br>20171 E<br>20171 E<br>20172 E<br>20173 E  | 6<br>6<br>6<br>6   |                     | populati              | on   |         |                                   |            |  |
| 23130695 (40)<br>22130655 (40)<br>23130695 (40)  | 20150 E<br>20150 E<br>20160 E<br>20170 E<br>20171 E<br>20172 E<br>20172 E<br>20173 E  | 6<br>6<br>6<br>6   |                     | 1.00                  |      |         |                                   |            |  |
| 22100455 140<br>21100459 140<br>2010055 140<br>20100455 140<br>20100455 140<br>20100455 140<br>20100455 140<br>20100455 140<br>20100455 140<br>20100455 140<br>20100459 140<br>20100459 140  | 29160 c<br>29169 c<br>29170 c<br>29171 c<br>29171 c<br>29172 c<br>29184 c<br>29186 c  | 6<br>6<br>6  |                     |                       |      |         |                                   |            |  |
| 23100898 140<br>23100858 140   | 201710 e<br>20171 e<br>20172 e<br>20174 e<br>20175 e  | ¢  |                     |                       |      |         |                                   |            |  |
| 22102655 1422<br>22102655 1422<br>22102655 1422<br>22102655 1422<br>22102655 1422<br>22102655 1422<br>22102655 1422<br>22102655 1422<br>22102655 1422  | 09.171 t<br>09.172 t<br>09.184 t<br>09.185 t  | ¢  |                     |                       |      |         | -6.2.4                            |            |  |
| 22130855 140<br>23130855 140<br>23130855 140<br>23130855 140<br>22130855 140<br>22130855 140<br>23130855 140<br>23130855 140<br>23130855 140   | 09.172 c<br>09.184 i<br>09.185 t  | ¢  |                     |                       |      | 493303  |                                   |            |  |
| 23138858 1437<br>23130858 1437<br>22130858 1437<br>22130858 1437<br>23138858 1437<br>23130858 1437<br>23130858 1437<br>23130858 1437   | 99.784 i<br>99.785 t  |  |                     |                       |      |         |                                   |            |  |
| 23130858 143<br>23130858 143<br>23130858 143<br>23130858 143<br>23130858 143<br>23130858 143<br>23130858 143   | 00.105 t  | ¢  |                     |                       |      | 1845    |                                   |            |  |
| 20100858 140<br>20100858 140<br>20100858 140<br>20100858 140<br>20100858 140   |   | ¢  |                     |                       |      | 種類あ     |                                   |            |  |
| 22120858 1420<br>23130858 1420<br>23130858 1420<br>23130858 1420<br>22120858 1420  |   |  |                     |                       |      | 1級職     |                                   |            |  |
| 23130858 140<br>23130858 140<br>22130858 140   |   | ¢  |                     |                       |      | 14844   |                                   |            |  |
| 23100858 140<br>22100858 140   |   | 9  |                     |                       |      | 機構      |                                   |            |  |
| 22100859 140   |   | 0  |                     |                       |      | 1818    |                                   |            |  |
|  |   | ¢  |                     |                       |      | 構想      |                                   |            |  |
|  |   | ¢  |                     |                       |      | 12038   |                                   |            |  |
| 23132848 142   |   | e<br>0   |                     |                       |      | 1933    |                                   |            |  |
| 23100858 142   |   | 0  |                     |                       |      | 1995    |                                   |            |  |
| 22105858 140   |   | c  |                     |                       |      | 報題      |                                   |            |  |
| 23105859 140   |   | 6  |                     |                       |      | 1833    |                                   |            |  |
| 23100858 142   |   | 0  |                     |                       |      | 1400007 | VG2 of: Gene 1                    | Name NOSZA |  |
| 22100858 140   |   | ¢  |                     |                       |      |         | JOINT MICH CR. BOOM               |            |  |
| 22100858 140   | 19.202 r  |  |                     |                       |      |         | mentrato conversantes             |            |  |
| 23108858 140   | 09.263 i  | 6  |                     |                       |      |         | fampyria<br>nature and            |            |  |
| 23100858 140   | 99.264 t  | ¢  |                     |                       |      |         | estin<br>es titals Pietestine - B |            |  |
| 22100858 140   | 1 2010  |  |                     |                       |      |         | outward by INDEDG                 |            |  |

| GVS: Genome V        | aria     | tio                         | n Se                                 | <b>IVEL BETA</b>  |   | nsored by SeattleSNPs<br>gram for Genemic Applications |
|----------------------|----------|-----------------------------|--------------------------------------|---|---|--|
| deglay<br>genetisses |          | displa<br>log snj           |                                      | y Results<br>disease interv                                   |   | SSOLY >  |
|                      | el Tress | nguane<br>nate 0.1          | y Cutoff (%)                         | ): 5<br>NEL, Soberitter: DSP_SNP                              |   |  |
|                      | 80       | Tatal<br>reunder<br>scittes | Average<br>Minur Atolio<br>Frequence | Tag SNP4  | CBer SNP4   |  |
|                      | ,        | 10                          | 12.%                                 | 231 38279   | 2211 3231<br>2211 3278<br>2211 475<br>2211 475<br>2211 475<br>2211 475<br>2212 455<br>2212 455<br>2212 455<br>2212 455<br>2214 4722 | //159489456492452235                                   |
|                      | 2        | ,                           | 35                                   | 231114572<br>23113441<br>23113537<br>23113111<br>23121290     | 20111443<br>20113093  |  |
| 14715309             | 3        | 5                           | 0.8                                  | 23125913  | 22108228<br>22118788<br>22112888<br>22112888<br>221141023   |  |
|                      |          | 4                           | 25%                                  | 234 12984<br>234 12982<br>234 19585<br>231 19585              |   |  |
|                      | 5        | +                           | 17.%                                 | 23123032<br>23124235<br>23124271<br>23124271<br>23125965      |   |  |
|                      |          | +                           | 24.%                                 | 22120412  | 22127181<br>22127214<br>22128748  |  |
|                      | 7        | +                           | 0.6                                  | 231 26899<br>231 40825<br>231 41732<br>231 41732<br>231 45040 |   | Madalian Calar sada                                    |
|                      |          | ¢                           | 11%                                  | 231 31 535  | 221120068<br>221120002<br>221120002<br>221120001  | Variation Color code<br>Coding                         |
|                      |          |                             |                                      |   |   |  |





| GVS: Gen   | ome         | Va                      | riati           | on Se         | rverm                      | га      | _                    | Sponsored by SeattleS |
|------------|-------------|-------------------------|-----------------|---------------|----------------------------|---------|----------------------|-----------------------|
| disp       | ay          | asplay<br>sop sumnary → |                 |               |                            |         |                      |                       |
| genot      |             |                         | 1.09            | snps →        | distre                     | QIIISAL | #)→                  |                       |
| Gene N     | ame: NO     | \$2A                    |                 |               |                            |         |                      |                       |
| Allele Fri | squency Cu  | ntoff (*                | ):5             |               |                            |         |                      |                       |
| Populatio  | IN: ECP_YC  | RUB-P.                  | MEL, Subr       | nitter: EGP_8 | NP8                        |         |                      |                       |
| 512 1000   | 81P /s ID   | ABBES                   | Property<br>(%) | Heteropyrety  | Hirdy-Weinberg<br>CH-Scewe | 08905   | function             |                       |
| 23108950   | 1011853718  | OLC.                    | 33              | 0.44          | 3.03                       | N062A   | Internet             |                       |
| 23108113   | /190088418  | TIC                     | 29              | 0.41          | 0.00                       | N0824   | internic             |                       |
| 23104235   | 197208775   | OV C                    | 18              | 0.23          | 0.23                       | N0824   | intranic             |                       |
| 22106897   | 1526544208  | NG                      | 12              | 0.23          | 0.22                       | N062A   | intranic             |                       |
| 22118756   | 1129944204  | TIC                     | 12              | 0.23          | 0.22                       | N0924   | intranic             |                       |
| 22111217   | 1120544201  | NO                      | 5               | 0.18          | 0.02                       | NOS24   | coding               |                       |
|            | 13729727    |                         | 36              | 0.46          | 0.35                       |         | intranic             |                       |
|            | rs3729551   | CFT                     | 41              | 0.48          | 1.10                       |         | Infranto             |                       |
|            | 192255920   | T/ A                    | 25              | 0.38          | 0.40                       |         | <b>Keanic</b>        |                       |
|            | 1\$3729507  | 0/ C                    | 18              | 0.23          | 0.23                       |         | intranic             |                       |
|            | 199068149   | 0/A                     | 26              | 0.38          | 0.16                       |         | intranic             |                       |
|            | 19906122    | NC                      | 17              | 0.28          | 0.47                       |         | intranic             |                       |
|            | 1020344158  |                         | 12              | 0.23          | 0.22                       |         | intranic<br>Intranic |                       |
|            | 1120544157  |                         | 17              | 0.28          | 0.47                       |         | intranic<br>Intranic |                       |
|            | 120844156   |                         | 25              | 0.31          | 4.26                       |         | intranic<br>Intranic |                       |
|            | 1920844195  |                         | 13              | 0.47          | *.20<br>0.23               |         | Parane:              |                       |
|            | 1930355831  | CIT                     | 42              | 0.49          | 1.65                       |         | KADING .             |                       |
|            | 1520844193  |                         | *2              | 0.47          | 0.71                       |         | internic             |                       |
|            | 10205441117 |                         | 28              | 0.41          | 0.00                       |         | intranic             |                       |
|            | 1125544152  |                         | 28              | 0.47          | 0.71                       |         | intranic             |                       |
|            | 1120944198  |                         | 29              | 0.41          | 0.00                       |         | Intranic             |                       |
|            | 11064826    |                         | 1               | 0.18          | 0.02                       |         | reding               |                       |
|            | 1120844102  |                         | 13              | 0.22          | 0.23                       |         | Infranto             |                       |
|            | 1928844181  |                         | 13              | 0.22          | 0.23                       |         | rear of              |                       |
|            |             | ne                      | - 10            |               | 0.00                       | -       |                      |                       |

# Finding SNPs: Databases and Extraction

One stop shopping - NIEHS SNPs and GeneSNPs

Prediction of functional variations - Polydoms and PolyPhen

Identification of trancription factor binding sites in Evolutionary Conserved Regions - TraFac and the ECR browser

Visualization and analysis of LD and TagSNPs - GVS