NCBI2R - To navigate and annotate genes and SNPs.
The Problem

• Genome Wide Analysis provides lots of p-values but without full context

The Solution

• Annotation
Why use NCBI 2 R?

• Simple to use

• No mirrors or out of date information

• Uses NCBI.

Scott Melville (ncbi2r@gmail.com)
GWA results – with SNP names, p-values, effect sizes etc

- GetSNPInfo
  - Position, Gene? Fxn_class, NCBI Locus ID numbers

- GetGeneInfo
  - Pathways, phenotypes, position, orientation, OMIM links, summaries, interacting genes

- GetNeighbours(genetic positions)
  - List of neighbouring genes within a user-specified distance (eg 100K). Creates links.

- MakeHTML
How To Analyse Results

<table>
<thead>
<tr>
<th>marker</th>
<th>p-values</th>
<th>n</th>
<th>beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs532523</td>
<td>0.015512</td>
<td>286</td>
<td>-0.5252</td>
</tr>
<tr>
<td>rs696786</td>
<td>0.000012</td>
<td>293</td>
<td>0.2142</td>
</tr>
<tr>
<td>rs626346</td>
<td>0.4240505</td>
<td>283</td>
<td>0.0622</td>
</tr>
</tbody>
</table>

> snplist<-anydf$marker
> snplist<-table[order(table$p),][1:100,"marker"]

> GetSNPInfo(snplist)
> GetSNPInfo("rs12456")

marker    genesymbol  locusID   chr   chrpos    fxn_class
rs12334    CIZ1        25792    9    129979750 missense

> GetSNPInfo(c("rs12456","rs626616"))

marker    genesymbol  locusID   chr   chrpos    fxn_class
rs12334    CIZ1        25792    9    129979750 missense
rs626616    19    60723974

Scott Melville (ncbi2r@gmail.com)
showurl=TRUE

> GetSNPInfo("rs12456",showurl=T)

GWA results – with SNP names, p-values, effect sizes etc

GetSNPInfo

Position, Gene? Fxn_class, NCBI Locus ID numbers

GetGeneInfo

Pathways, phenotypes, position, orientation, OMIM links, summaries, interacting genes

GetNeighbours(genetic positions)

List of neighbouring genes within a user-specified distance (eg 100K). Creates links.

makeHTML

Scott Melville (ncbi2r@gmail.com)
marker     gene     locusID   chr   chrpos      fxn_class
rs12334    CIZ1     25792     9     129979750  missense

1: CIZ1-CDKN1A interacting zinc finger protein 1 [Homo sapiens]
GeneID: 25792

Official Symbol: CIZ1
Official Full Name: CDKN1A interacting zinc finger protein 1
Primary source: HGNC:16744
See related: Ensembl:ENSG00000148337; HPRD:13061; MIM:611420
Gene type: protein coding
RefSeq status: VALIDATED
Organism: Homo sapiens
Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as: NP94; LSFR1; ZNF356; CIZ1
>GetGeneInfo(25792)

locusID 25792
Org_ref_taxname/comm name Homo sapiens, human
OMIM 611420
synonyms NP94 LSFR1 ZNF356
genesummary
genename CDKN1A interacting zinc finger protein 1
phenotypes pathways
GenePos 129968165 130006483
Ori -
Chromosome 9
genesymbol CIZ1
Interim 0
The centromere is a specialized chromatin domain, present throughout the cell cycle, that acts as a platform on which the transient assembly of the kinetochore occurs during mitosis. All active centromeres are characterized by the presence of long arrays of nucleosomes in which CENPA (MIM 117139) replaces histone H3 (see MIM 601128). CENPN is an additional factor required for centromere assembly (Foltz et al., 2006 [PubMed 16622419]). [supplied by OMIM]

genename phenotypes
centromere protein N

pathways
Reactome Event: Cell Cycle, Mitotic
GWA results – with SNP names, p-values, effect sizes etc

GetSNPInfo

GetGeneInfo

Position, Gene? Fxn_class, NCBI Locus ID numbers

Pathways, phenotypes, position, orientation, OMIM links, summaries, interacting genes

GetNeighbours(genetic positions)

List of neighbouring genes within a user-specified distance (eg 100K). Creates links.

makeHTML

Scott Melville (ncbi2r@gmail.com)
GWA results – with SNP names, p-values, effect sizes etc

GetSNPInfo
  Position, Gene? Fxn_class, NCBI Locus ID numbers

GetGeneInfo
  Pathways, phenotypes, position, orientation, OMIM links, summaries, interacting genes

GetNeighbours(genetic positions)
  List of neighbouring genes within a user-specified distance (eg 100K). Creates links.

makeHTML
<table>
<thead>
<tr>
<th>marker</th>
<th>chr</th>
<th>chrpos</th>
<th>fn_class</th>
<th>genesymbol</th>
<th>genename</th>
<th>OMIM</th>
<th>synonyms</th>
<th>genesummary</th>
<th>phenotypes</th>
<th>pathways</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs626262</td>
<td>6</td>
<td>101836492</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>This gene encodes a member of the alpha-1 subunit family*, a protein in the voltage-dependent calcium channel complex. Calcium channels mediate the influx of calcium ions into the cell upon membrane polarization and consist of a complex of alpha-1, alpha-2/delta, beta, and gamma subunits in a 1:1:1:1 ratio. The alpha-1 subunit has 24 transmembrane segments and forms the pore through which ions pass into the cell. There are multiple isoforms of each of the proteins in the complex, either encoded by different genes or the result of alternative splicing of transcripts. Alternate transcriptional splice variants of the gene described here have been observed but have not been thoroughly characterized. Mutations in this gene have been associated with some neurological disorders.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs61653156</td>
<td>X</td>
<td>48948950</td>
<td>intron</td>
<td>CACNA1F</td>
<td></td>
<td></td>
<td></td>
<td>Aland Island eye disease --- Cone-rod dystrophy, X-linked. 3 --- Night blindness, congenital stationary, X-linked type 2 --- Night blindness, congenital stationary, X-linked type 2A --- Ocular albinism, Forsius-Eriksson type</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Shows SNP info**

**Opens up pubmed references**

**Opens a visual map**

**Shows Gene info**
<table>
<thead>
<tr>
<th>synonyms</th>
<th>genesummary</th>
<th>phenotypes</th>
<th>pathways</th>
<th>GeneStartPos</th>
<th>GeneStopPos</th>
<th>Ori</th>
<th>Neighbours</th>
</tr>
</thead>
<tbody>
<tr>
<td>This gene encodes a member of the alpha-1 subunit family*, a protein in the voltage-dependent calcium channel complex. Calcium channels mediate the influx of calcium ions into the cell upon membrane polarization and consist of a complex of alpha-1, alpha-2/delta, beta, and gamma subunits in a 1:1:1:1 ratio. The alpha-1 subunit has 24 transmembrane segments and forms the pore through which ions pass into the cell. There are multiple isoforms of each of the proteins in the complex, either encoded by different genes or the result of alternative splicing of transcripts. Alternate transcriptional splice variants of the gene described here have been observed but have not been thoroughly characterized. Mutations in this gene have</td>
<td>Aland Island eye disease---Cone-rod dystrophy, X-linked, 3---Night blindness, congenital stationary, X-linked, type 2---Night blindness, congenital stationary, X-linked, type 2A---Ocular albinism, Forsius-Eriksson type</td>
<td>--- KEGG pathway: Calcium signaling pathway---KEGG pathway: GnRH signaling pathway---KEGG pathway: MAPK signaling pathway</td>
<td>48948466</td>
<td>48976776</td>
<td>-</td>
<td>synaptophysin**, forkhead box P3**, calcium channel, voltage-dependent, L type, alpha 1F subunit**, proteolipid protein 2 (colonic epithelium-enriched)<strong>, prickle homolog 3 (Drosophila)</strong>, coiled-coil domain containing 22<strong>G patch domain and KOW motifs</strong>, MAGI family member, X-linked<strong>G antigen 10</strong>, protein phosphatase 1, regulatory (inhibitor) subunit 3F**, heat shock 27kDa protein-like 3, pseudogene**, hypothetical protein LOC643767</td>
<td></td>
</tr>
</tbody>
</table>
AnnotateDataframe
(mydata, selections=c(“marker”,”p”,”beta”),
filename=“bone_results.html”)

Similar Functions:
AnnotateSNPlist, AnnotateSNPfile
ScanForGenes

GWA results

GetID("ENST004142")

GetID("sleep[DIS]")

GetID("protein binding[GO]")

GetID("CLN5[sym]")

GetID("CLN5")

GetID("KEGG pathway: Cytokine-cytokine receptor interaction")

(from literature)
<table>
<thead>
<tr>
<th></th>
<th>Gene</th>
<th>Chr</th>
<th>Start</th>
<th>Stop</th>
<th>Orientation</th>
<th>StartFlank</th>
<th>StopFlank</th>
<th>bestSNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23560</td>
<td>10</td>
<td>1024348</td>
<td>1053707</td>
<td>plus</td>
<td>924348</td>
<td>1055707</td>
<td>rs1904671</td>
</tr>
<tr>
<td></td>
<td>bestP</td>
<td>nSNPs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1028181</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Gene</th>
<th>Chr</th>
<th>Start</th>
<th>Stop</th>
<th>Orientation</th>
<th>SNPname</th>
<th>pos</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs7069505</td>
<td>23560</td>
<td>10</td>
<td>1024348</td>
<td>plus</td>
<td>rs7069505</td>
<td>946400</td>
<td>0.1523807</td>
</tr>
<tr>
<td>1</td>
<td>rs10904588</td>
<td>23560</td>
<td>10</td>
<td>1024348</td>
<td>plus</td>
<td>rs10904588</td>
<td>981737</td>
<td>0.9849032</td>
</tr>
<tr>
<td>1</td>
<td>rs6560861</td>
<td>23560</td>
<td>10</td>
<td>1024348</td>
<td>plus</td>
<td>rs6560861</td>
<td>984753</td>
<td>0.5459408</td>
</tr>
<tr>
<td>1</td>
<td>rs1904671</td>
<td>23560</td>
<td>10</td>
<td>1024348</td>
<td>plus</td>
<td>rs1904671</td>
<td>1004603</td>
<td>0.1028181</td>
</tr>
<tr>
<td>1</td>
<td>rs947403</td>
<td>23560</td>
<td>10</td>
<td>1024348</td>
<td>plus</td>
<td>rs947403</td>
<td>1006601</td>
<td>0.2450576</td>
</tr>
<tr>
<td>1</td>
<td>rs4880745</td>
<td>23560</td>
<td>10</td>
<td>1024348</td>
<td>plus</td>
<td>rs4880745</td>
<td>1008585</td>
<td>0.2701996</td>
</tr>
<tr>
<td>1</td>
<td>rs7069123</td>
<td>23560</td>
<td>10</td>
<td>1024348</td>
<td>plus</td>
<td>rs7069123</td>
<td>1014214</td>
<td>0.3111333</td>
</tr>
<tr>
<td>1</td>
<td>rs7920922</td>
<td>23560</td>
<td>10</td>
<td>1024348</td>
<td>plus</td>
<td>rs7920922</td>
<td>1018179</td>
<td>0.5749135</td>
</tr>
</tbody>
</table>
ScanForSNPs

Candidates list

GWA results

from literature

for replication

```r
> ScanForSNPs(c("rs7532643","rs4757589","rs6134143"),msf_corr_poly_meta)
```

<table>
<thead>
<tr>
<th>ReqSNP</th>
<th>ReqChr</th>
<th>ReqPos</th>
<th>distance</th>
<th>name</th>
<th>chromosome</th>
<th>position</th>
<th>strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7532643</td>
<td>1</td>
<td>198925807</td>
<td>670</td>
<td>rs4915451</td>
<td>1</td>
<td>198926477</td>
<td>+</td>
</tr>
<tr>
<td>rs4757589</td>
<td>11</td>
<td>17767535</td>
<td>1235</td>
<td>rs1528</td>
<td>3</td>
<td>17766300</td>
<td>+</td>
</tr>
<tr>
<td>rs6134143</td>
<td>20</td>
<td>1116773</td>
<td>1218</td>
<td>rs6108855</td>
<td>13</td>
<td>1117991</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>allele1</th>
<th>allele2</th>
<th>effallele</th>
<th>effallelefreq</th>
<th>n</th>
<th>beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>G</td>
<td>G</td>
<td>0.432253313696613</td>
<td>679</td>
<td>0.0786414950318363</td>
</tr>
<tr>
<td>T</td>
<td>C</td>
<td>C</td>
<td>0.0922401171303075</td>
<td>683</td>
<td>0.049612404993172</td>
</tr>
<tr>
<td>A</td>
<td>G</td>
<td>G</td>
<td>0.092375366568915</td>
<td>682</td>
<td>-0.0743044555400157</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sebeta</th>
<th>p</th>
<th>pgc</th>
<th>pexhwe</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0487407351970893</td>
<td>0.106643411230825</td>
<td>0.10817687924656</td>
<td>0.481854167340122</td>
</tr>
<tr>
<td>0.0809523931773222</td>
<td>0.539969556615518</td>
<td>0.541735205094724</td>
<td>0.499066129181896</td>
</tr>
<tr>
<td>0.0803794063565681</td>
<td>0.355266878171132</td>
<td>0.357365148245419</td>
<td>0.355373031747563</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>call</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0.998535871156662</td>
</tr>
</tbody>
</table>
refs<-GetPubMed("CLN5",download=TRUE)

MakeExcel(refs,"References.tab")

OpenPMID(18371232)
OpenPDF(18371232)
A mutation in canine CLN5 causes neuronal ceroid lipofuscinosis in Border collie dogs.

Melville SA, Wilson CL, Chiang CS, Studdert VP, Lingaas F, Wilton AN.

School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW 2052, Australia.

Neuronal ceroid lipofuscinosis (NCL) is a neurodegenerative disease found in Border collie dogs, humans, and other animals. Disease gene studies in humans and animals provided candidates for the NCL gene in Border collie dogs.
GetRegion

- GetRegion("snp","4",12300000,24100000)
- GetRegion("gene","X",624642,984642)
Other Functions

• GetIDs(“CLN5[sym]”)

• GetGeneTable(1203)
  – exons, introns, transcripts

• GetGOs(1203)

• GetInteractions(1203)
Other Functions

• GetPathways(1203)

• GetPhenotypes(1203)

• GetSNPsInGene(1203)

• And nothing to do with NCBI…
• NatureJobs(c("genetics","statistics"))