NCBI2R - To navigate and annotate genes and SNPs.

The Problem

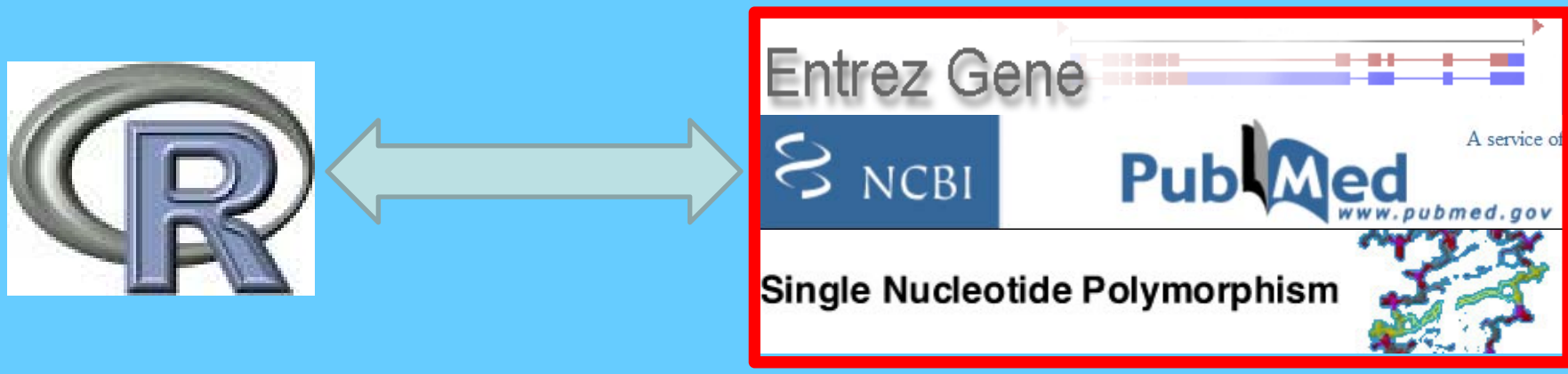
- Genome Wide Analysis provides lots of p-values but without full context

The Solution

- Annotation

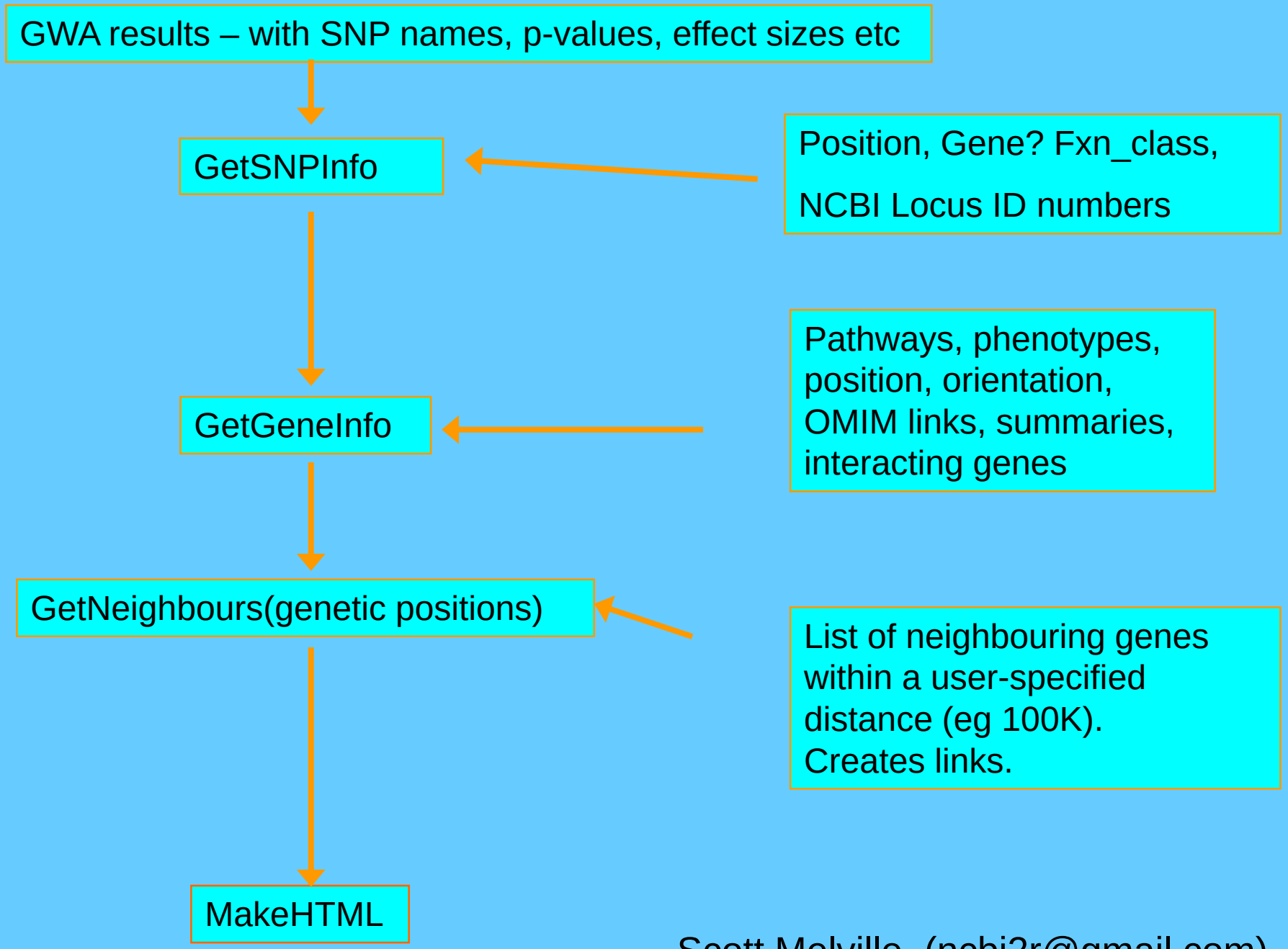
Scott Melville (ncbi2r@gmail.com)

Why use NCBI 2 R?



- Simple to use
- No mirrors or out of date information
- Uses NCBI.

Scott Melville (ncbi2r@gmail.com)



Scott Melville (ncbi2r@gmail.com)

How To Analyse Results

marker	p-values	n	beta
rs532523	0.015512	286	-0.5252
rs696786	0.000012	293	0.2142
rs626346	0.4240505	283	0.0622

> snplist<-anydf$marker

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

> GetSNPInfo(snplist)

Scott Melville (ncbi2r@gmail.com)

GetSNPInfo

> GetSNPInfo(“rs12456”)

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

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

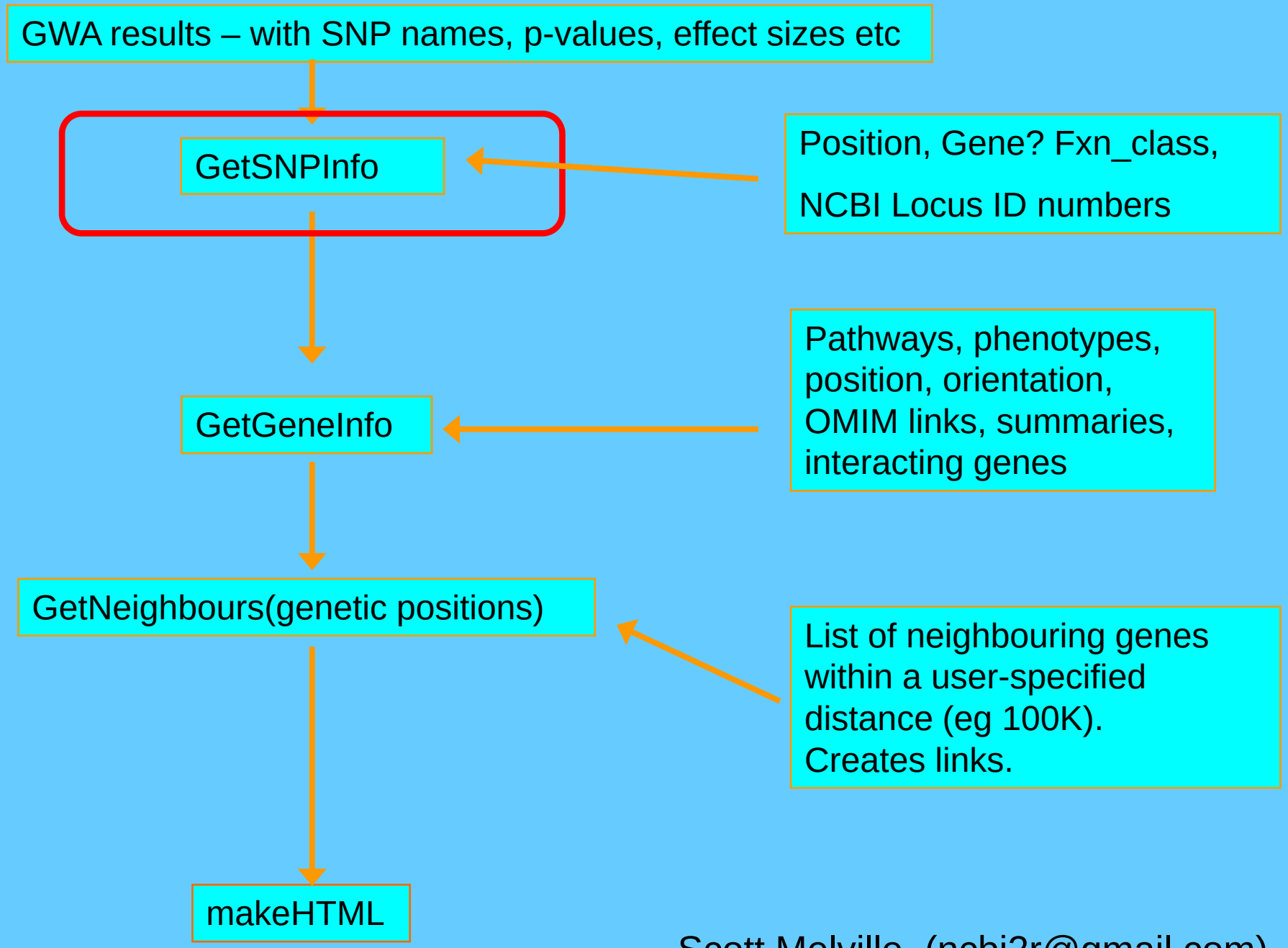
marker	genesymbol	locusID	chr	chrpos	fxn_class
rs12334	CIZ1	25792	9	129979750	missense
rs626616			19	60723974	

Scott Melville (ncbi2r@gmail.com)

showurl=TRUE

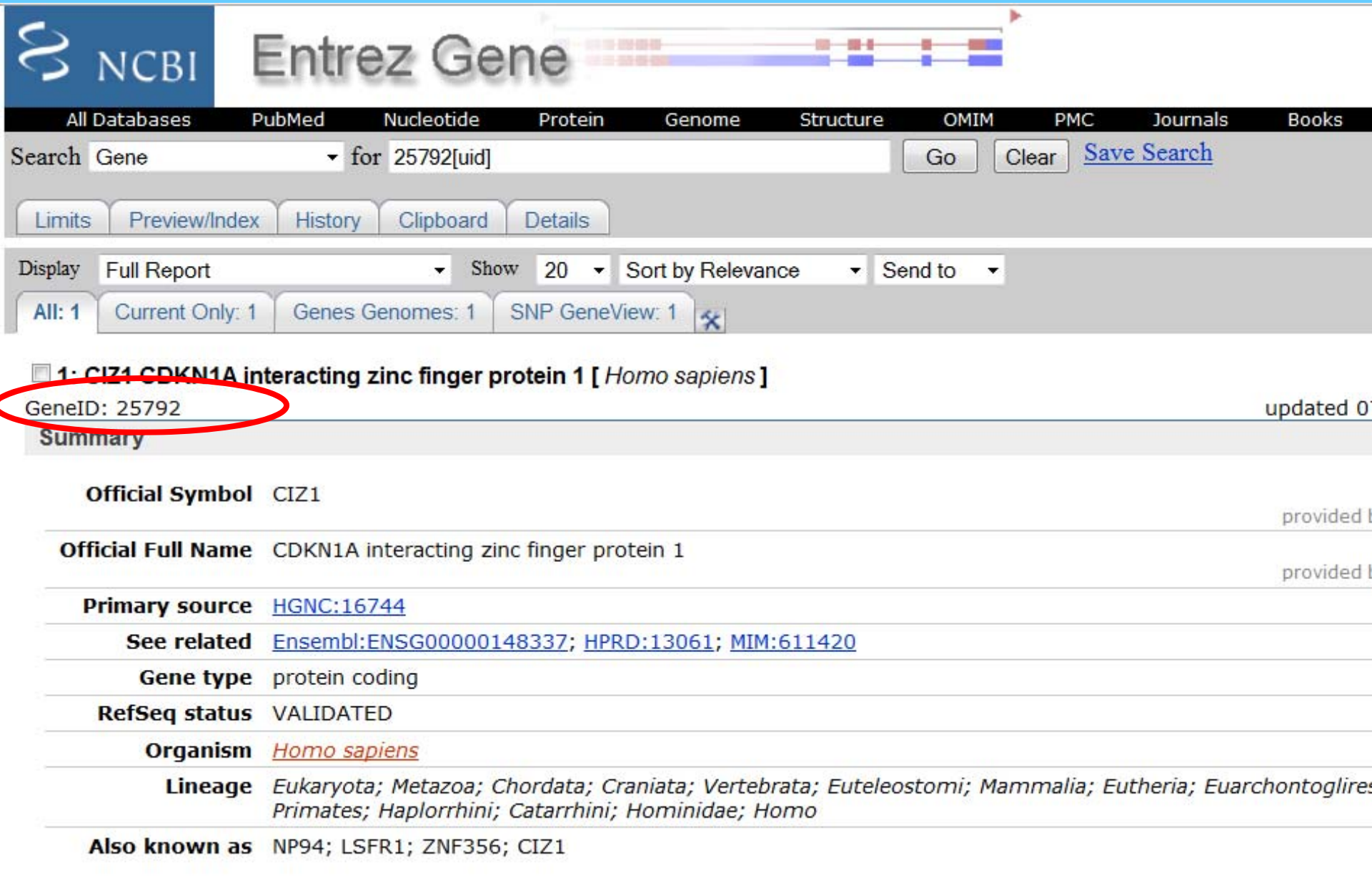
> GetSNPInfo(“rs12456”,showurl=T)

http://eutils.ncbi.nlm.nih.gov/entrez/eutils/
efetch.fcgi?db=snp&id=12356
&report=DocSet
&tool=NCBI2R&email=ncbi2r@gmail.com



Scott Melville (ncbi2r@gmail.com)

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



>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

>**GetGeneInfo(55839)**

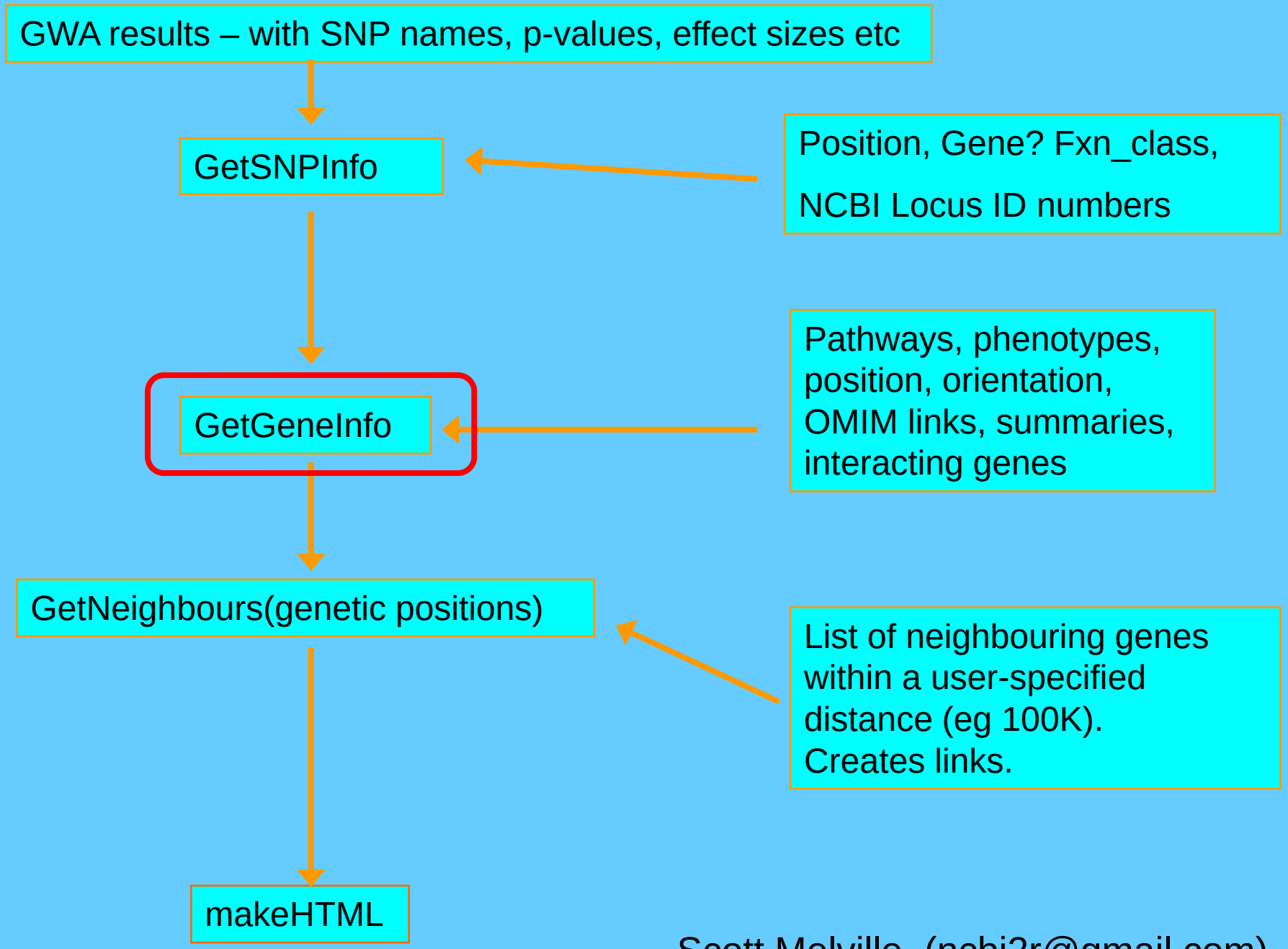
locusID	OMIM	synonyms
55839	611509	BM039 CENP-N C16orf60 FLJ13607 FLJ22660

genesummary
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]

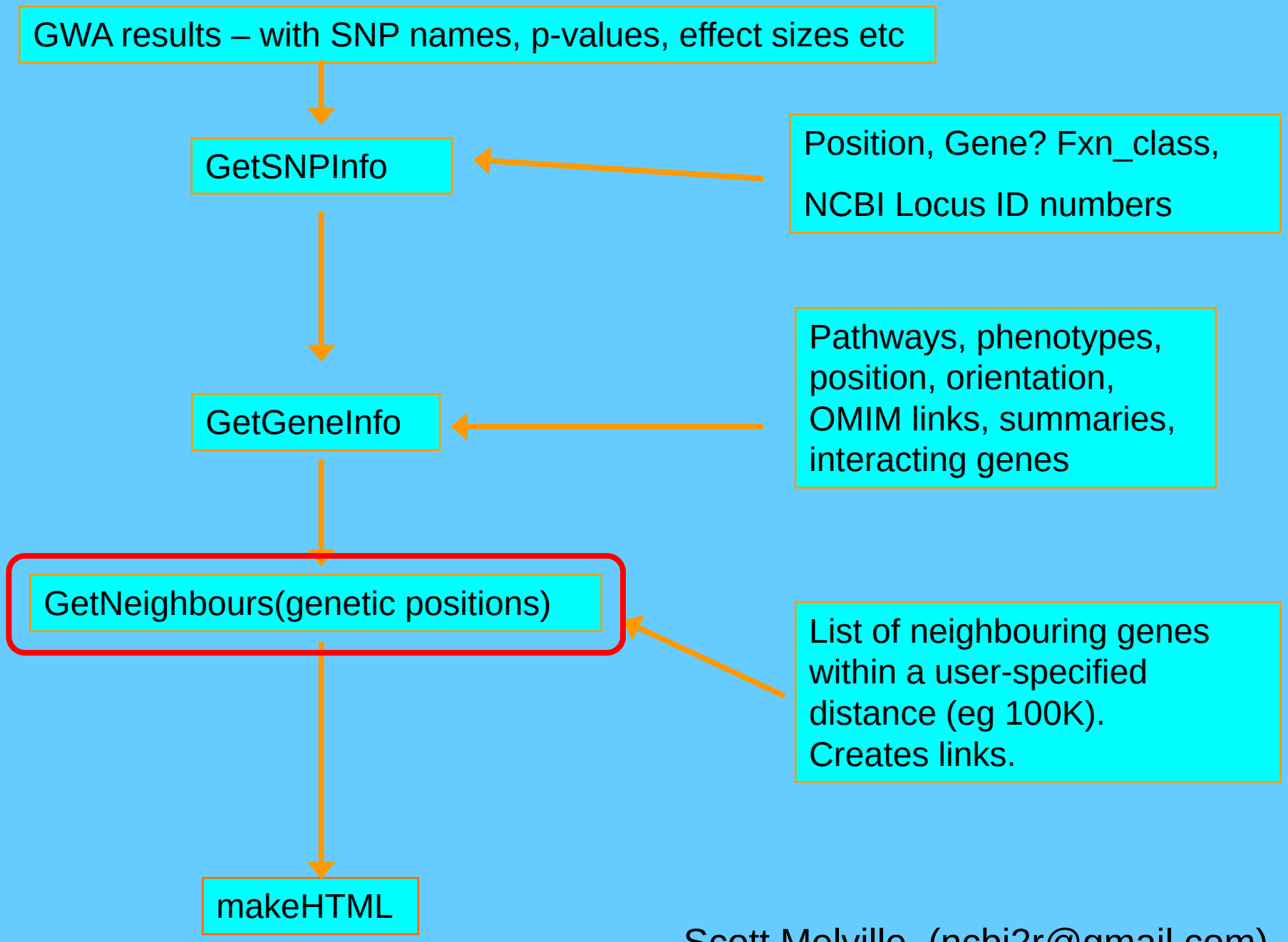
genename	phenotypes
centromere protein N	

pathways
Reactome Event:Cell Cycle, Mitotic

GeneStartPos	GeneStopPos	Ori	Chromosome	genesymbol
79597603	79624212	+	16	CENPN



Scott Melville (ncbi2r@gmail.com)



Scott Melville (ncbi2r@gmail.com)

marker	chr	chrpos	fxn_class	genesymbol	genename	OMIM	synonyms	genesummary	phenotypes	pathways
rs626262	6	101836492								
rs61653156	X	48948950	intron	CACNA1F	calcium channel, voltage-dependent, L type, alpha 1F subunit	OMIM	JM8 OA2 AIED COD3 JMC8 CORDX CSNB2 CORDX3 CSNB2A CSNBX2 Cav1.4	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.	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	---KEGG pathway: Calcium signaling pathway---KEGG pathway: GnRH signaling pathway---KEGG pathway: MAPK signaling pathway

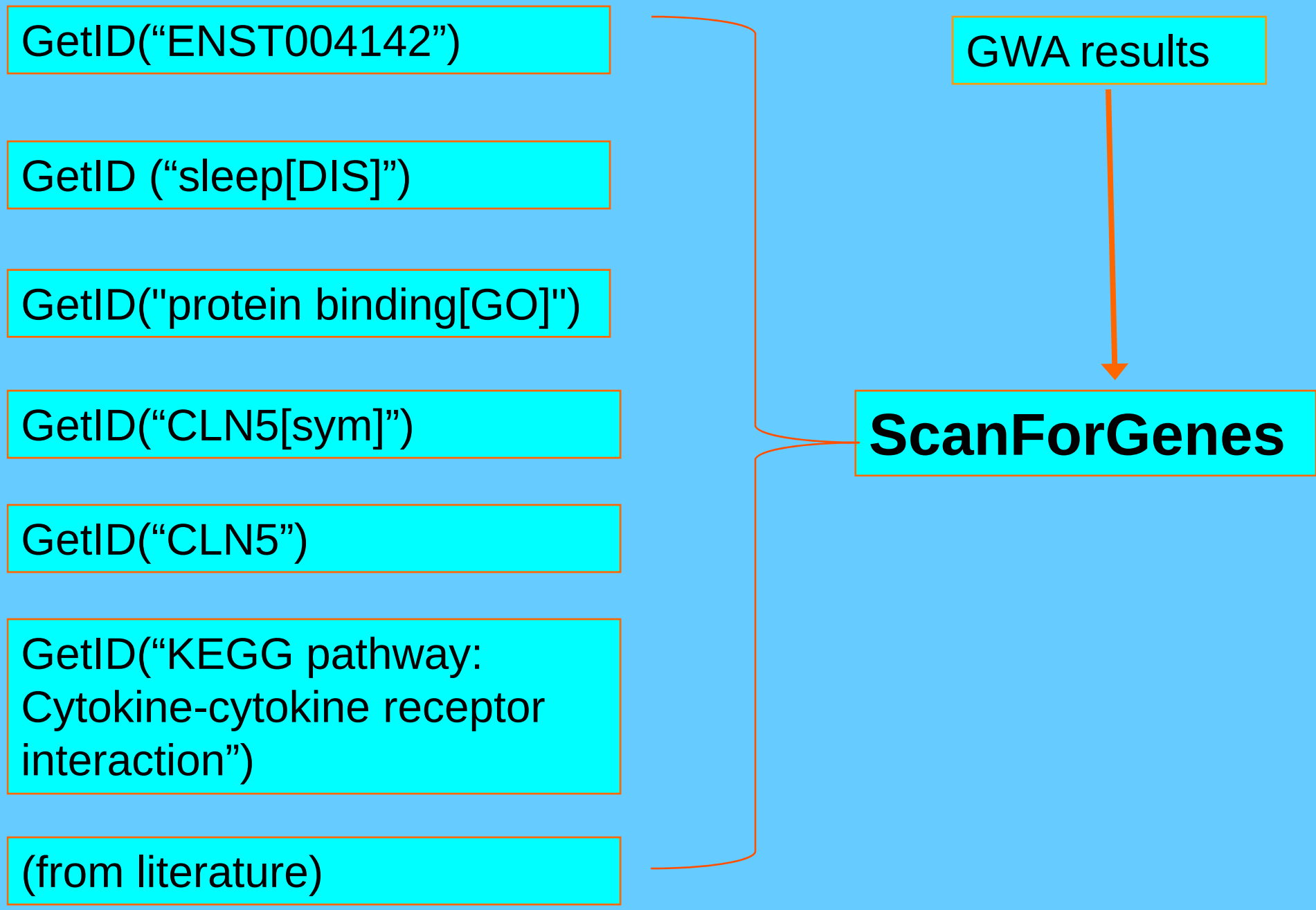
synonyms	genesummary	phenotypes	pathways	GeneStartPos	GeneStopPos	Ori	Neighbours
JM8 OA2 AIED COD3 JMC8 CORDX CSNB2 CORDX3 CSNB2A CSNBX2 Cav1.4	This gene encodes a member of the alpha-1 subunit family*, a protein in the voltage-dependent calcium channel complex. Calcium channels mediate the influx of 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.	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	---KEGG pathway: Calcium signaling pathway---KEGG pathway: GnRH signaling pathway---KEGG pathway: MAPK signaling pathway	48948466	48976776	-	synaptophysin**forkhead box P3**calcium channel, voltage-dependent, L type, alpha 1F subunit**proteolipid protein 2 (colonic epithelium-enriched)**prickle homolog 3 (Drosophila)**coiled-coil domain containing 22**G patch domain and KOW motifs**MAGI family member, X-linked**G antigen 10**protein phosphatase 1, regulatory (inhibitor) subunit 3F**heat shock 27kDa protein-like 3 pseudogene**hypothetical protein LOC643767

EasyFunctions

AnnotateDataframe
(mydata,selections=c(“marker”,”p”,”beta”),
filename=“bone_results.html”)

Similar Functions:
AnnotateSNPlist, AnnotateSNPfile

Scott Melville (ncbi2r@gmail.com)


GetID(“ENST004142”)
GetID (“sleep[DIS]”)
GetID("protein binding[GO]")
GetID(“CLN5[sym]”)
GetID(“CLN5”)
GetID(“KEGG pathway: Cytokine-cytokine receptor interaction”)
(from literature)
GWA results
ScanForGenes

ScanForGenes

```
$best
   Gene Chr   Start    Stop Orientation StartFlank StopFlank   bestSNP
1 23560  10 1024348 1053707        plus     924348   1055707 rs1904671
      bestP nSNPs
1 0.1028181    13

$complete
            Gene Chr   Start    Stop Orientation    SNPname    pos      p
rs7069505   23560  10 1024348 1053707        plus  rs7069505  946400 0.1523807
rs10904588 23560  10 1024348 1053707        plus rs10904588  981737 0.9849032
rs6560861   23560  10 1024348 1053707        plus  rs6560861  984753 0.5459408
rs1904671   23560  10 1024348 1053707        plus  rs1904671 1004603 0.1028181
rs947403    23560  10 1024348 1053707        plus   rs947403 1006601 0.2450576
rs4880745   23560  10 1024348 1053707        plus  rs4880745 1008585 0.2701996
rs7069123   23560  10 1024348 1053707        plus  rs7069123 1014214 0.3111333
rs7920922   23560  10 1024348 1053707        plus  rs7920922 1018179 0.5749135
```

Position	*SNPnames*	*p−values*
62166322	rs6062342	7.4E-01
62183120	rs6090041	1.5E-01
62198359	rs7271530	1.0E-01
62199875	rs2229205	4.4E-01
62208012	rs4809401	2.3E-01
62210011	rs910883	6.1E-01
62213766	rs3003155	5.7E-01
62221898	rs13039817	1.3E-01
62233330	rs530703	9.6E-01
62243081	rs2983435	9.3E-01
62252526	rs4809256	9.5E-01
62256817	rs4809258	2.7E-01
62257372	rs3003172	2.8E-01
62260916	rs508396	7.8E-01
62265999	rs379042	8.8E-01
62271625	rs576479	2.0E-01
62288203	rs16984240	1.7E-01
62296587	rs2012987	9.7E-01
62298952	rs13038037	5.7E-01
62301246	rs3003137	2.7E-01
62306964	rs6062662	1.6E-01
62316387	rs2983460	3.1E-01
62316975	rs912239	2.5E-01
62321464	rs6122274	7.3E-01
62330452	rs2427621	9.0E-01
62331424	rs2427622	6.2E-01

File created at 2009−03−16 17:25:07

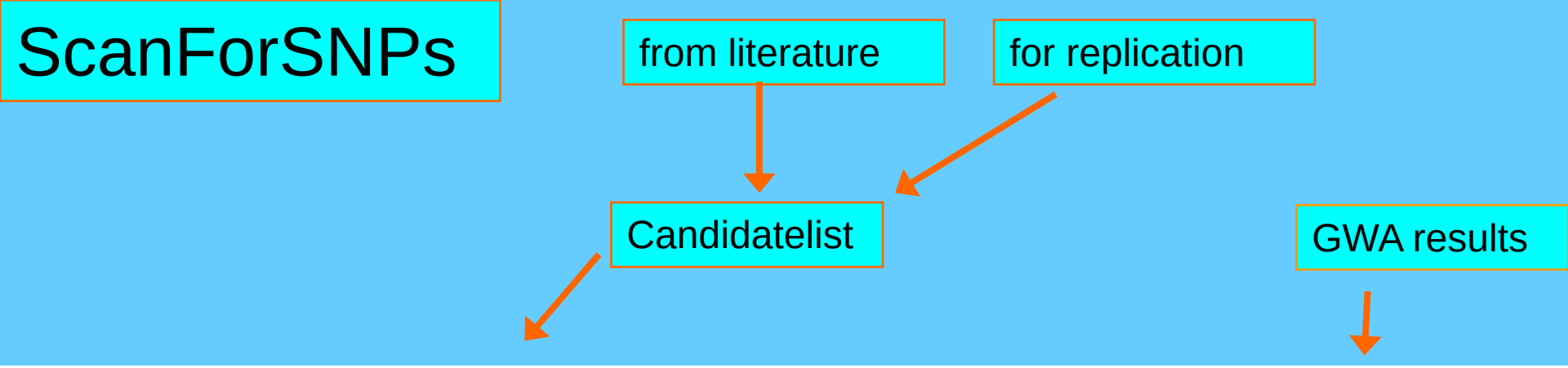
Chromosome 20

myelin transcription factor 1[4661]

Sig Level

Transcript (Exons,Introns)

ScanForSNPs




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

	ReqSNP	ReqChr	ReqPos	distance	name	chromosome	position	strand
1	rs7532643	1	198925807	670	rs4915451	1	198926477	+
2	rs4757589	11	17767535	1235	rs1528	3	17766300	+
3	rs6134143	20	1116773	1218	rs6108855	13	1117991	+

	allele1	allele2	effallele	effallelefreq	n	beta
1	T	G	G	0.432253313696613	679	0.0786414950318363
2	T	C	C	0.092240117130307 5	683	0.049612404993172
3	A	G	G	0.092375366568915	682	-0.0743044555400157

	sebeta	p	pgc	pexhwe
1	0.0487407351970893	0.106643411230825	0.108176879924656	0.481854167340122
2	0.0809523931773222	0.539969556615518	0.541735205094724	0.499066129181896
3	0.0803794063565681	0.355266878171132	0.357365148245419	0.355373031747563

	call
1	0.994143484626647
2	1
3	0.998535871156662

refs<-GetPubMed("CLN5",download=TRUE)

MakeExcel(refs,"References.tab")

OpenPMID(18371232)

OpenPDF(18371232)

Scott Melville (ncbi2r@gmail.com)

1: Genomics. 2005 Sep;86(3):287-94.

ELSEVIER FULL-TEXT ARTICLE Lin

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

Related articles

- Neuronal ceroid lipofuscinosis in Devon cattle is caused by a single b: [Biochim Biophys Acta. 200
- CLN5, a novel gene encoding a putative transmembrane protein mutated i [Nat Genet. 199
- A CLN5 mutation causing an atypical neuronal ceroid lipofuscinosis of juvenile or [Neurology. 200
- Review Mutated genes in juvenile and variant lat infantile neuronal ceroid lipofus [Curr Mol Med. 200

GetRegion

- GetRegion(“snp”,”4”,12300000,24100000)
- GetRegion(“gene”,”X”,624642,984642)

Scott Melville (ncbi2r@gmail.com)

Other Functions

- GetIDs(“CLN5[sym]”)
- GetGeneTable(1203)
  - exons, introns, transcripts
- GetGOs(1203)
- GetInteractions(1203)

Scott Melville (ncbi2r@gmail.com)

Other Functions

- GetPathways(1203)
- GetPhenotypes(1203)
- GetSNPsInGene(1203)
- And nothing to do with NCBI…
- NatureJobs(c(“genetics”,”statistics”))

Scott Melville (ncbi2r@gmail.com)