rsnps tutorial

Install and load library

When available on CRAN

```
install.packages("rsnps")
```

Or get from Github

```
install.packages("devtools")
library(devtools)
install_github("rsnps", "ropensci")
```

```
library(rsnps)
```

Get genotype data for all users at a particular snp.

```
allgensnp(snp = "rs7412")[1:3]
```

```
[[1]]
[[1]]$snp
[[1]]$snp$name
[1] "rs7412"

[[1]]$snp$chromosome
[1] "19"

[[1]]$snp$position
[1] "50103919"


[[1]]$user
[[1]]$user$name
[1] "Lisa"

[[1]]$user$id
[1] 1653

[[1]]$user$genotypes
[[1]]$user$genotypes[[1]]
[[1]]$user$genotypes[[1]]$genotype_id
[1] 944

[[1]]$user$genotypes[[1]]$local_genotype
[1] "CC"
```

```
[[2]]
[[2]]$snp
[[2]]$snp$name
[1] "rs7412"

[[2]]$snp$chromosome
[1] "19"

[[2]]$snp$position
[1] "50103919"


[[2]]$user
[[2]]$user$name
[1] "karl"

[[2]]$user$id
[1] 1651

[[2]]$user$genotypes
[[2]]$user$genotypes[[1]]
[[2]]$user$genotypes[[1]]$genotype_id
[1] 943

[[2]]$user$genotypes[[1]]$local_genotype
[1] "CC"




[[3]]
[[3]]$snp
[[3]]$snp$name
[1] "rs7412"

[[3]]$snp$chromosome
[1] "19"

[[3]]$snp$position
[1] "50103919"


[[3]]$user
[[3]]$user$name
[1] "bpaslc"

[[3]]$user$id
[1] 1639

[[3]]$user$genotypes
[[3]]$user$genotypes[[1]]
[[3]]$user$genotypes[[1]]$genotype_id
```

```
[1] 933

[[3]]$user$genotypes[[1]]$local_genotype
[1] "CT"
```

```
allgensnp("rs7412", df = TRUE)[1:10, ]
```

```
   snp_name snp_chromosome snp_position       user_name user_id
1    rs7412             19     50103919            Lisa    1653
2    rs7412             19     50103919            karl    1651
3    rs7412             19     50103919          bpaslc    1639
4    rs7412             19     50103919         Wally97    1641
5    rs7412             19     50103919            Paul    1635
6    rs7412             19     50103919          Arthur    1621
7    rs7412             19     50103919 Justin Anzalone    1620
8    rs7412             19     50103919    Brenda Ramos    1619
9    rs7412             19     50103919 Jeremy McEntire    1617
10   rs7412             19     50103919        jonathan    1616
   genotype_id genotype
1          944       CC
2          943       CC
3          933       CT
4          935       CT
5          931       CC
6          919       CC
7          918       CC
8          917       CC
9          915       CC
10         914       CT
```

Get all phenotypes, their variations, and how many users have data available for a given phenotype.

Get all data

```
allphenotypes(df = TRUE)[1:10, ]
```

```
   id characteristic                    known_variations number_of_users
1   1      Eye color                               Brown             411
2   1      Eye color                          Brown-green             411
3   1      Eye color                           Blue-green             411
4   1      Eye color                            Blue-grey             411
5   1      Eye color                                Green             411
6   1      Eye color                                 Blue             411
7   1      Eye color                               Hazel             411
8   1      Eye color                               Mixed             411
9   1      Eye color                           Gray-blue             411
10  1      Eye color Blue-grey; broken amber collarette             411
```

Output a list, then call the characterisitc of interest by 'id' or 'characteristic'

```r
datalist <- allphenotypes()
names(datalist)[1:10]  # get list of all characteristics you can call
```

```
 [1] "Eye color"           "Handedness"          "Height"
 [4] "Sex"                 "Hair Color"          "Tongue roller"
 [7] "Colour Blindness"    "Lactose intolerance" "white skin"
[10] "Coffee consumption"
```

```r
datalist[["ADHD"]]  # get data.frame for 'ADHD'
```

```
  id characteristic                              known_variations
1 29           ADHD                                         False
2 29           ADHD                                          True
3 29           ADHD              Undiagnosed, but probably true
4 29           ADHD                                            No
5 29           ADHD                                           Yes
6 29           ADHD                                  Not diagnosed
7 29           ADHD Diagnosed as not having but with some signs
8 29           ADHD                                    Mthfr c677t
  number_of_users
1             114
2             114
3             114
4             114
5             114
6             114
7             114
8             114
```

```r
datalist[c("mouth size", "SAT Writing")]  # get data.frame for 'ADHD'
```

```
$`mouth size`
   id characteristic known_variations number_of_users
1 120     mouth size           Medium              44
2 120     mouth size            Small              44
3 120     mouth size            Large              44

$`SAT Writing`
   id characteristic                known_variations number_of_users
1  41    SAT Writing                             750              37
2  41    SAT Writing              Tested before 2005              37
3  41    SAT Writing                             800              37
4  41    SAT Writing             Country with no sat              37
5  41    SAT Writing                             N/a              37
6  41    SAT Writing         Never & have ba & above              37
7  41    SAT Writing                             720              37
8  41    SAT Writing                             511              37
9  41    SAT Writing Did well - don't remember score              37
10 41    SAT Writing                             700              37
```

Get annotations for a given snp.

Get just the metadata

```
annotations(snp = "rs7903146", output = "metadata")
```

```
         .id        V1
1       name rs7903146
2 chromosome        10
3   position 114748339
```

Just from PLOS journals

```
annotations(snp = "rs7903146", output = "plos")[c(1:10), ]
```

```
                author
1  Marguerite R. Irvin
2         Huixiao Hong
3         Daniel Savic
4  Jeanne M. McCaffery
5        Cornelia Then
6      Changzheng Dong
7    Anette P. Gjesing
8  Jeanne M. McCaffery
9          Jinjin Wang
10      Jingxiang Chen

1                 Genome-Wide Detection of Allele Specific Copy Number Variation Associated with Insulin Re
2                                                    Technical Reproducibility of Genotyping SNP Arrays Us
3                   An <i>in vivo cis</i>-Regulatory Screen at the Type 2 Diabetes Associated <i>TCF7L2</i>
4                                           <i>TCF7L2</i> Polymorphism, Weight Loss and Proinsulin Insuli
5  Plasma Metabolomics Reveal Alterations of Sphingo- and Glycerophospholipid Levels in Non-Diabetic Carrier
6                                                                 Gene-Centric Characteristics of G
7                       The Effect of <i>PCSK1</i> Variants on Waist, Waist-Hip Ratio and Glucose Metaboli
8                                           <i>TCF7L2</i> Polymorphism, Weight Loss and Proinsulin Insuli
9                 Association of rs7903146 (IVS3C/T) and rs290487 (IVS3C/T) Polymorphisms in <i>TCF7L2</i>
10                                                    Association between TCF7L2 Gene Polymorphis
       publication_date number_of_readers
1  2011-08-25T00:00:00Z              1427
2  2012-09-07T00:00:00Z               509
3  2012-05-10T00:00:00Z               697
4  2011-07-26T00:00:00Z              1421
5  2013-10-24T00:00:00Z              none
6  2007-12-05T00:00:00Z              none
7  2011-09-14T00:00:00Z               296
8  2011-07-26T00:00:00Z              1421
9  2013-03-25T00:00:00Z              none
10 2013-08-09T00:00:00Z              none
                                          url
1  http://dx.doi.org/10.1371/journal.pone.0024052
2  http://dx.doi.org/10.1371/journal.pone.0044483
3  http://dx.doi.org/10.1371/journal.pone.0036501
4  http://dx.doi.org/10.1371/journal.pone.0021518
```

```
5  http://dx.doi.org/10.1371/journal.pone.0078430
6  http://dx.doi.org/10.1371/journal.pone.0001262
7  http://dx.doi.org/10.1371/journal.pone.0023907
8  http://dx.doi.org/10.1371/journal.pone.0021518
9  http://dx.doi.org/10.1371/journal.pone.0059053
10 http://dx.doi.org/10.1371/journal.pone.0071730
                           doi
1  10.1371/journal.pone.0024052
2  10.1371/journal.pone.0044483
3  10.1371/journal.pone.0036501
4  10.1371/journal.pone.0021518
5  10.1371/journal.pone.0078430
6  10.1371/journal.pone.0001262
7  10.1371/journal.pone.0023907
8  10.1371/journal.pone.0021518
9  10.1371/journal.pone.0059053
10 10.1371/journal.pone.0071730
```

Just from SNPedia

```
annotations(snp = "rs7903146", output = "snpedia")
```

```
                                              url
1 http://www.snpedia.com/index.php/Rs7903146(C;C)
2 http://www.snpedia.com/index.php/Rs7903146(C;T)
3 http://www.snpedia.com/index.php/Rs7903146(T;T)
                                                                summary
1 Normal (lower) risk of Type 2 Diabetes and Gestational Diabetes.
2     1.4x increased risk for diabetes (and perhaps colon cancer).
3                           2x increased risk for Type-2 diabetes
```

Get all annotations

```
annotations(snp = "rs7903146", output = "all")[1:10, ]
```

```
        .id                      author
1  mendeley        Dhanasekaran Bodhini
2  mendeley Ludmila Alves Sanches Dutra
3  mendeley               Thomas Hansen
4  mendeley    Laura J Rasmussen-Torvik
5  mendeley                     Yu Yan
6  mendeley                  K Pilgaard
7  mendeley       André Gustavo P Sousa
8  mendeley             Stéphane Cauchi
9  mendeley    Panagiotis Christopoulos
10 mendeley           Martha L Slattery

1                                                The rs12255372(G/T) and rs7903146(C/T) polym
2                                                     Allele-specific PCR assay to ge
3                                                                          At-Risk
4                                                  Preliminary report: No association between
5                                The transcription factor 7-like 2 (TCF7L2) polymorphism may be associate
```

```
6  The T allele of rs7903146 TCF7L2 is associated with impaired insulinotropic action of incretin hormones, r
7                                                                                                  TCF7L2 Polymorphis
8                                                                                               TCF7L2 rs7903146 varian
9                                                                                                   Genetic variants in
10
   publication_year number_of_readers open_access
1              2007                 8       FALSE
2              2008                 5       FALSE
3              2011                 1       FALSE
4              2009                 3       FALSE
5              2010                 5        TRUE
6              2009                 8       FALSE
7              2009                11        TRUE
8              2007                 4        TRUE
9              2008                 2       FALSE
10             2008                 4       FALSE

1                                                                 http://www.mendeley.com/research/rs12255372-g
2                                                               http://www.mendeley.com/research/allelespecific-pcr-
3                                                                                    http://www.mendeley.co
4                                                             http://www.mendeley.com/research/preliminary-report-a
5                                                   http://www.mendeley.com/research/transcription-factor-7like-2-tcf
6  http://www.mendeley.com/research/t-allele-rs7903146-tcf7l2-associated-impaired-insulinotropic-action-
7                                                                          http://www.mendeley.com/research/
8                                                                 http://www.mendeley.com/research/tcf7l2-rs7903
9                                                                       http://www.mendeley.com/resear
10                                                                           http://www.mer
                           doi publication_date summary first_author
1                         none             <NA>    <NA>         <NA>
2                         none             <NA>    <NA>         <NA>
3  10.1016/j.biopsych.2011.01.031             <NA>    <NA>         <NA>
4                         none             <NA>    <NA>         <NA>
5          10.1186/1472-6823-10-9             <NA>    <NA>         <NA>
6                         none             <NA>    <NA>         <NA>
7    10.1371/journal.pone.0007697             <NA>    <NA>         <NA>
8          10.1186/1471-2350-8-37             <NA>    <NA>         <NA>
9                         none             <NA>    <NA>         <NA>
10                        none             <NA>    <NA>         <NA>
   pubmed_link journal trait pvalue pvalue_description confidence_interval
1         <NA>    <NA>  <NA>     NA               <NA>                <NA>
2         <NA>    <NA>  <NA>     NA               <NA>                <NA>
3         <NA>    <NA>  <NA>     NA               <NA>                <NA>
4         <NA>    <NA>  <NA>     NA               <NA>                <NA>
5         <NA>    <NA>  <NA>     NA               <NA>                <NA>
6         <NA>    <NA>  <NA>     NA               <NA>                <NA>
7         <NA>    <NA>  <NA>     NA               <NA>                <NA>
8         <NA>    <NA>  <NA>     NA               <NA>                <NA>
9         <NA>    <NA>  <NA>     NA               <NA>                <NA>
10        <NA>    <NA>  <NA>     NA               <NA>                <NA>
```

Download genotype data for a user from 23andme or other repo.

```
data <- users(df = TRUE)
head(data[[1]])  # users with links to genome data
fetch_genotypes(url = data[[1]][1, "genotypes.download_url"], rows = 15)
```

Get genotype data for one or multiple users.

```
genotypes(snp = "rs9939609", userid = 1)
```

```
$snp
$snp$name
[1] "rs9939609"

$snp$chromosome
[1] "16"

$snp$position
[1] "52378028"


$user
$user$name
[1] "Bastian Greshake"

$user$id
[1] 1

$user$genotypes
$user$genotypes[[1]]
$user$genotypes[[1]]$genotype_id
[1] 9

$user$genotypes[[1]]$local_genotype
[1] "AT"
```

```
genotypes("rs9939609", userid = "1,6,8", df = TRUE)
```

```
   snp_name snp_chromosome snp_position        user_name user_id
1 rs9939609             16     52378028 Bastian Greshake       1
2 rs9939609             16     52378028     Nash Parovoz       6
3 rs9939609             16     52378028         Samantha       8
  genotype_id genotype
1           9       AT
2           5       AT
3           2       TT
```

```
genotypes("rs9939609", userid = "1-2", df = FALSE)
```

```
[[1]]
[[1]]$snp
[[1]]$snp$name
[1] "rs9939609"

[[1]]$snp$chromosome
[1] "16"

[[1]]$snp$position
[1] "52378028"


[[1]]$user
[[1]]$user$name
[1] "Bastian Greshake"

[[1]]$user$id
[1] 1

[[1]]$user$genotypes
[[1]]$user$genotypes[[1]]
[[1]]$user$genotypes[[1]]$genotype_id
[1] 9

[[1]]$user$genotypes[[1]]$local_genotype
[1] "AT"




[[2]]
[[2]]$snp
[[2]]$snp$name
[1] "rs9939609"

[[2]]$snp$chromosome
[1] "16"

[[2]]$snp$position
[1] "52378028"


[[2]]$user
[[2]]$user$name
[1] "Senficon"

[[2]]$user$id
[1] 2

[[2]]$user$genotypes
list()
```

Get phenotype data for one or multiple users.

```
phenotypes(userid = 1)$phenotypes[1:3]
```

```
$`white skin`
$`white skin`$phenotype_id
[1] 4

$`white skin`$variation
[1] "Caucasian"


$`Lactose intolerance`
$`Lactose intolerance`$phenotype_id
[1] 2

$`Lactose intolerance`$variation
[1] "lactose-tolerant"


$`Eye color`
$`Eye color`$phenotype_id
[1] 1

$`Eye color`$variation
[1] "blue-green"
```

```
phenotypes(userid = "1,6,8", df = TRUE)[[1]][1:10, ]
```

```
                   phenotype phenotypeID                 variation
1                 white skin           4                 Caucasian
2        Lactose intolerance           2           lactose-tolerant
3                  Eye color           1                 blue-green
4                  Hair Type          16                   straight
5                     Height          15            Tall ( >180cm )
6               Ability to Tan          14                        Yes
7  Short-sightedness (Myopia)          21                        low
8        Nicotine dependence           20 Smoker. 10 cigarettes/day
9                Beard Color           12                      Blonde
10          Colour Blindness           25                       False
```

```
out <- phenotypes(userid = "1-8", df = TRUE)
lapply(out, head)
```

```
$`Bastian Greshake`
            phenotype phenotypeID        variation
1          white skin           4        Caucasian
2 Lactose intolerance           2 lactose-tolerant
3           Eye color           1       blue-green
4           Hair Type          16         straight
5              Height          15  Tall ( >180cm )
```

```
6      Ability to Tan          14              Yes

$Senficon
  phenotype phenotypeID variation
1   no data     no data   no data

$`no info on user_3`
  phenotype phenotypeID variation
1   no data     no data   no data

$`no info on user_4`
  phenotype phenotypeID variation
1   no data     no data   no data

$`no info on user_5`
  phenotype phenotypeID variation
1   no data     no data   no data

$`Nash Parovoz`
                        phenotype phenotypeID        variation
1                       Handedness           3     right-handed
2                        Eye color           1            brown
3                       white skin           4        Caucasian
4              Lactose intolerance           2 lactose-tolerant
5 Ability to find a bug in openSNP           5   extremely high
6           Number of wisdom teeth          57                4

$`no info on user_7`
  phenotype phenotypeID variation
1   no data     no data   no data

$Samantha
                   phenotype phenotypeID                   variation
1 Short-sightedness (Myopia)          21                      medium
2                 Handedness           3                left-handed
3        Lactose intolerance           2          lactose-intolerant
4                  Eye color           1                       Brown
5             Ability to Tan          14                         Yes
6        Nicotine dependence          20 ex-smoker, 7 cigarettes/day
```

Get all known variations and all users sharing that phenotype for one phenotype(-ID).

```
phenotypes_byid(phenotypeid = 12, return_ = "desc")
```

```
$id
[1] 12

$characteristic
[1] "Beard Color"

$description
[1] "coloration of facial hair"
```

```
phenotypes_byid(phenotypeid = 12, return_ = "knownvars")
```

```
$known_variations
 [1] "Red"
 [2] "Blonde"
 [3] "Red-brown"
 [4] "Red-blonde-brown-black(in diferent parts i have different color,for example near the lips blond-red"
 [5] "No beard-female"
 [6] "Brown-black"
 [7] "Blonde-brown"
 [8] "Black"
 [9] "Dark brown with minor blondish-red"
[10] "Brown-grey"
[11] "Red-blonde-brown-black"
[12] "Blond-brown"
[13] "Brown, some red"
[14] "Brown"
[15] "Brown-gray"
[16] "Never had a beard"
[17] "I'm a woman"
[18] "Black-brown-blonde"
[19] "Was red-brown now mixed with gray,"
[20] "Red-blonde-brown"
```

```
phenotypes_byid(phenotypeid = 12, return_ = "users")[1:10, ]
```

```
   user_id
1       22
2        1
3       26
4       10
5       14
6       42
7       45
8       16
9        8
10     661
                                                                                variation
1                                                                                     Red
2                                                                                  Blonde
3                                                                                red-brown
4  Red-Blonde-Brown-Black(in diferent parts i have different color,for example near the lips blond-red
5                                                                          No beard-female
6                                                                              Brown-black
7  Red-Blonde-Brown-Black(in diferent parts i have different color,for example near the lips blond-red
8                                                                            blonde-brown
9                                                                          No beard-female
10                                                                             Brown-black
```

Get openSNP users.

```
data <- users(df = FALSE)
data[1:2]
```

```
[[1]]
[[1]]$name
[1] "gigatwo"

[[1]]$id
[1] 31

[[1]]$genotypes
list()


[[2]]
[[2]]$name
[1] "Anu Acharya"

[[2]]$id
[1] 385

[[2]]$genotypes
list()
```

Search for SNPs in Linkage Disequilibrium with a set of SNPs

```
LDSearch("rs420358")
```

```
Querying SNAP...
Querying NCBI for up-to-date SNP annotation information...
Done!

$rs420358
       Proxy      SNP Distance RSquared DPrime GeneVariant GeneName
4   rs420358 rs420358        0    1.000  1.000  INTERGENIC      N/A
5   rs442418 rs420358      122    1.000  1.000  INTERGENIC      N/A
8   rs718223 rs420358     1168    1.000  1.000  INTERGENIC      N/A
6   rs453604 rs420358     2947    1.000  1.000  INTERGENIC      N/A
3   rs372946 rs420358      -70    0.943  1.000  INTERGENIC      N/A
1 rs10889290 rs420358     3987    0.800  1.000  INTERGENIC      N/A
2 rs10889291 rs420358     4334    0.800  1.000  INTERGENIC      N/A
7  rs4660403 rs420358     7021    0.800  1.000  INTERGENIC      N/A
  GeneDescription Major Minor   MAF NObserved Chromosome_NCBI Marker_NCBI
4             N/A     C     A 0.167       120               1    rs420358
5             N/A     C     T 0.167       120               1    rs442418
8             N/A     A     G 0.167       120               1    rs718223
6             N/A     A     G 0.167       120               1    rs453604
3             N/A     G     C 0.175       120               1    rs372946
```

```
1             N/A     G     A 0.200       120               1  rs10889290
2             N/A     C     T 0.200       120               1  rs10889291
7             N/A     A     G 0.200       120               1   rs4660403
  Class_NCBI Gene_NCBI Alleles_NCBI Major_NCBI Minor_NCBI MAF_NCBI
4        snp      <NA>          G/T          G          T   0.0891
5        snp      <NA>          A/G          G          A   0.0891
8        snp      <NA>          A/G          A          G   0.0891
6        snp      <NA>          A/G          A          G   0.0836
3        snp      <NA>          C/G          G          C   0.0891
1        snp      <NA>          A/G          G          A   0.1015
2        snp      <NA>          C/T          C          T   0.1015
7        snp      <NA>          A/G          A          G   0.0969
   BP_NCBI
4 40806910
5 40807032
8 40808078
6 40809857
3 40806840
1 40810897
2 40811244
7 40813931
```

Query NCBI's dbSNP for information on a set of SNPs

An example with both merged SNPs, non-SNV SNPs, regular SNPs, SNPs not found, microsatellite

```
snps <- c("rs332", "rs420358", "rs1837253", "rs1209415715", "rs111068718")
NCBI_snp_query(snps)
```

```
        Query Chromosome      Marker          Class Gene   Alleles Major
1       rs332          7 rs121909001         in-del CFTR     -/TTT  <NA>
2    rs420358          1    rs420358            snp <NA>       G/T     G
3   rs1837253          5   rs1837253            snp <NA>       C/T     C
4 rs111068718       <NA> rs111068718 microsatellite <NA> (GT)21/24  <NA>
  Minor    MAF        BP
1  <NA>     NA 117199646
2     T 0.0891  40806910
3     T 0.3627 110401871
4  <NA>     NA        NA
```