# Anàlisi funcional de les dades de càncer de pulmó

Xavier Duran Albareda
GCAT Genomes for Life
Institut de Recerca Germans Trias i Pujol (IGTP)

HealthForecast
13 de Desembre del 2016

# Gene Set Analysis (GSA)

# Variant list

| name | chrom | chromStart | chromEnd | stability | scaled_coefficients | ref | alt |
|---|---|---|---|---|---|---|---|
| 12:65995405:C:G | 12 | 65995405 | 65995406 | 10 | 0.6901523 | C | G |
| 12:132695710:G:A | 12 | 132695710 | 132695711 | 10 | 0.6184025 | G | A |
| 12:122307580:A:G | 12 | 122307580 | 122307581 | 6 | 0.5574568 | A | G |
| 12:86433258:A:C | 12 | 86433258 | 86433259 | 8 | 0.5441592 | C | A |
| 12:63058957:A:G | 12 | 63058957 | 63058958 | 10 | 0.5285514 | G | A |
| 12:121149596:A:G | 12 | 121149596 | 121149597 | 6 | 0.5188642 | G | A |
| 12:127933561:C:CT | 12 | 127933561 | 127933562 | 10 | 0.5086084 | C | CT |
| 12:26860830:CA:C | 12 | 26860830 | 26860831 | 9 | 0.4982072 | C | CA |
| 12:64966695:T:C | 12 | 64966695 | 64966696 | 10 | 0.4958828 | C | T |
| 12:78538314:C:T | 12 | 78538314 | 78538315 | 8 | 0.4950422 | T | C |
| 12:127933756:A:G | 12 | 127933756 | 127933757 | 10 | 0.4940085 | A | G |
| 12:118388554:A:T | 12 | 118388554 | 118388555 | 8 | 0.4931533 | T | A |
| 12:2011473:C:A | 12 | 2011473 | 2011474 | 4 | 0.4925469 | C | A |
| 12:614204:A:G | 12 | 614204 | 614205 | 8 | 0.4921972 | G | A |
| 12:52954624:C:A | 12 | 52954624 | 52954625 | 6 | 0.4888366 | A | C |
| 12:64963697:G:A | 12 | 64963697 | 64963698 | 9 | 0.4887630 | A | G |
| 12:91262911:T:C | 12 | 91262911 | 91262912 | 8 | 0.4886846 | C | T |
| 12:117985667:C:T | 12 | 117985667 | 117985668 | 8 | 0.4884988 | T | C |
| 12:18075974:G:A | 12 | 18075974 | 18075975 | 3 | 0.4883262 | A | G |
| 12:125500901:G:GT | 12 | 125500901 | 125500902 | 8 | 0.4879617 | GT | G |

*Taula 1: Llista de variants*

# From variants to genes

## Ensembl

Provide a centralized resource for geneticists, molecular biologists and other researchers studying the genomes of our own species and other vertebrates and model organisms.

One of several well known browsers for the retrieval of genomic information.

# From variants to genes

## Chromosome 12

| | |
|---|---|
| Length (bps) | 133,851,895 |
| Coding genes | 1,071 |
| Non coding gene count | 1,131 |
| Pseudogenes | 616 |
| Short Variants | 6,980,191 |

# From variants to genes

## Seq2pathway

Seq2pathway is an R/Python wrapper for pathway (or functional gene-set) analysis of genomic loci, adapted for advances in genome research.

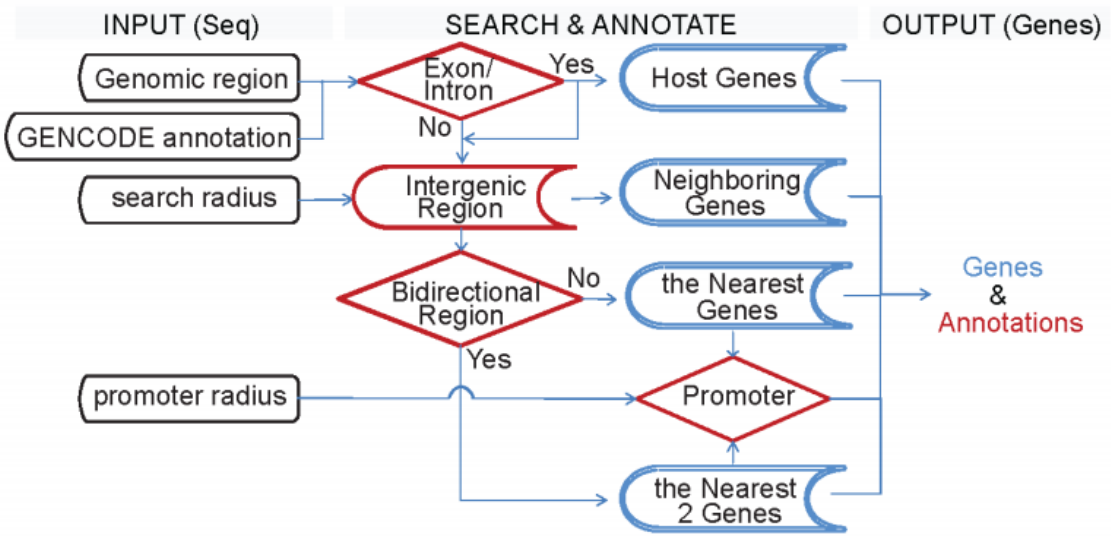



*Figura 1: seq2gene workflow*

# From variants to genes

## Code

```
gene_dat <- runseq2gene(
  inputfile = snp_dat,
  genome = "hg19",
  adjacent = TRUE,
  SNP = TRUE,
  search_radius = 150000,
  PromoterStop = FALSE,
  NearestTwoDirection = TRUE
)
```

# From variants to genes

## Variants and nearby genes

| name | chrom | chromStart | chromEnd | type | gene_name | source |
|---|---|---|---|---|---|---|
| 12:13500475:A:G | 12 | 13500475 | 13500476 | Nearest_R | C12orf36 | protein_coding |
| 12:106248532:A:G | 12 | 106248532 | 106248533 | Nearest_R | NUAK1 | protein_coding |
| 12:84616756:A:G | 12 | 84616756 | 84616757 | Nearest_R | SLC6A15 | protein_coding |
| 12:87452755:T:C | 12 | 87452755 | 87452756 | Nearest_R | C12orf29 | protein_coding |
| 12:127506324:C:T | 12 | 127506324 | 127506325 | Intron | RP11-575F12.1 | lincRNA |
| 12:71363764:T:C | 12 | 71363764 | 71363765 | Nearest_L | PTPRR | protein_coding |
| 12:68319817:T:A | 12 | 68319817 | 68319818 | Nearest_L | DYRK2 | protein_coding |
| 12:10307972:A:G | 12 | 10307972 | 10307973 | Nearest_L | CLEC7A | protein_coding |
| rs10735363 | 12 | 99524896 | 99524897 | Intron | ANKS1B | protein_coding |
| 12:58093835:A:G | 12 | 58093835 | 58093836 | Intron | OS9 | protein_coding |
| 12:94979538:A:G | 12 | 94979538 | 94979539 | Intron | TMCC3 | protein_coding |
| rs2521979 | 12 | 47814159 | 47814160 | Nearest_R | RPAP3 | protein_coding |
| 12:67006088:A:G | 12 | 67006088 | 67006089 | Intron | GRIP1 | protein_coding |
| exm981054 | 12 | 7894056 | 7894057 | Exon | CLEC4C | protein_coding |
| 12:94944787:G:A | 12 | 94944787 | 94944788 | Nearest_R | TMCC3 | protein_coding |
| 12:114729723:A:C | 12 | 114729723 | 114729724 | Nearest_R | TBX5 | protein_coding |
| 12:99262556:G:A | 12 | 99262556 | 99262557 | Intron | ANKS1B | protein_coding |
| 12:14864819:GA:G | 12 | 14864819 | 14864820 | Nearest_R | HIST4H4 | protein_coding |
| 12:31471585:GA:G | 12 | 31471585 | 31471586 | Intron | FAM60A | protein_coding |
| 12:53132528:G:A | 12 | 53132528 | 53132529 | Nearest_R | KRT76 | protein_coding |

*Taula 3: Llista de variants*

# From variants to genes

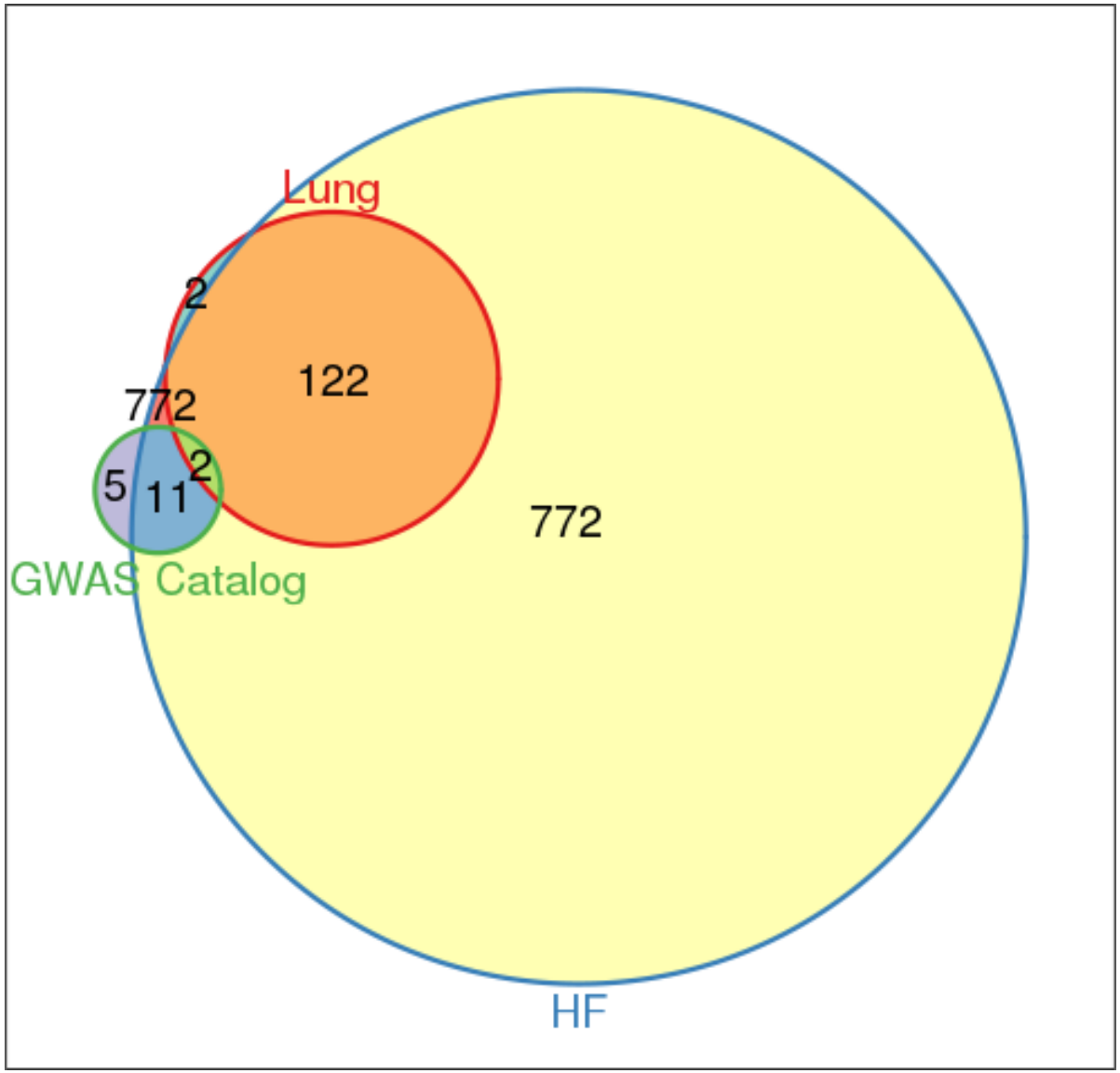
## Gene list

| gene | frequency |
|---|---|
| CCDC91 | 360 |
| TMEM132B | 345 |
| TMEM132C | 340 |
| ATF7IP | 333 |
| IFLTD1 | 326 |
| ANKS1B | 303 |
| GRIP1 | 265 |
| AEBP2 | 247 |
| PPFIA2 | 247 |
| NEDD1 | 229 |
| RP11-181C3.1 | 214 |
| SOX5 | 188 |
| NAV3 | 161 |
| AC092850.1 | 160 |
| C12orf79 | 157 |
| TMEM132D | 156 |
| CMKLR1 | 155 |
| DIP2B | 148 |
| SFSWAP | 148 |
| CRADD | 141 |

*Taula 4: Llista de variants*

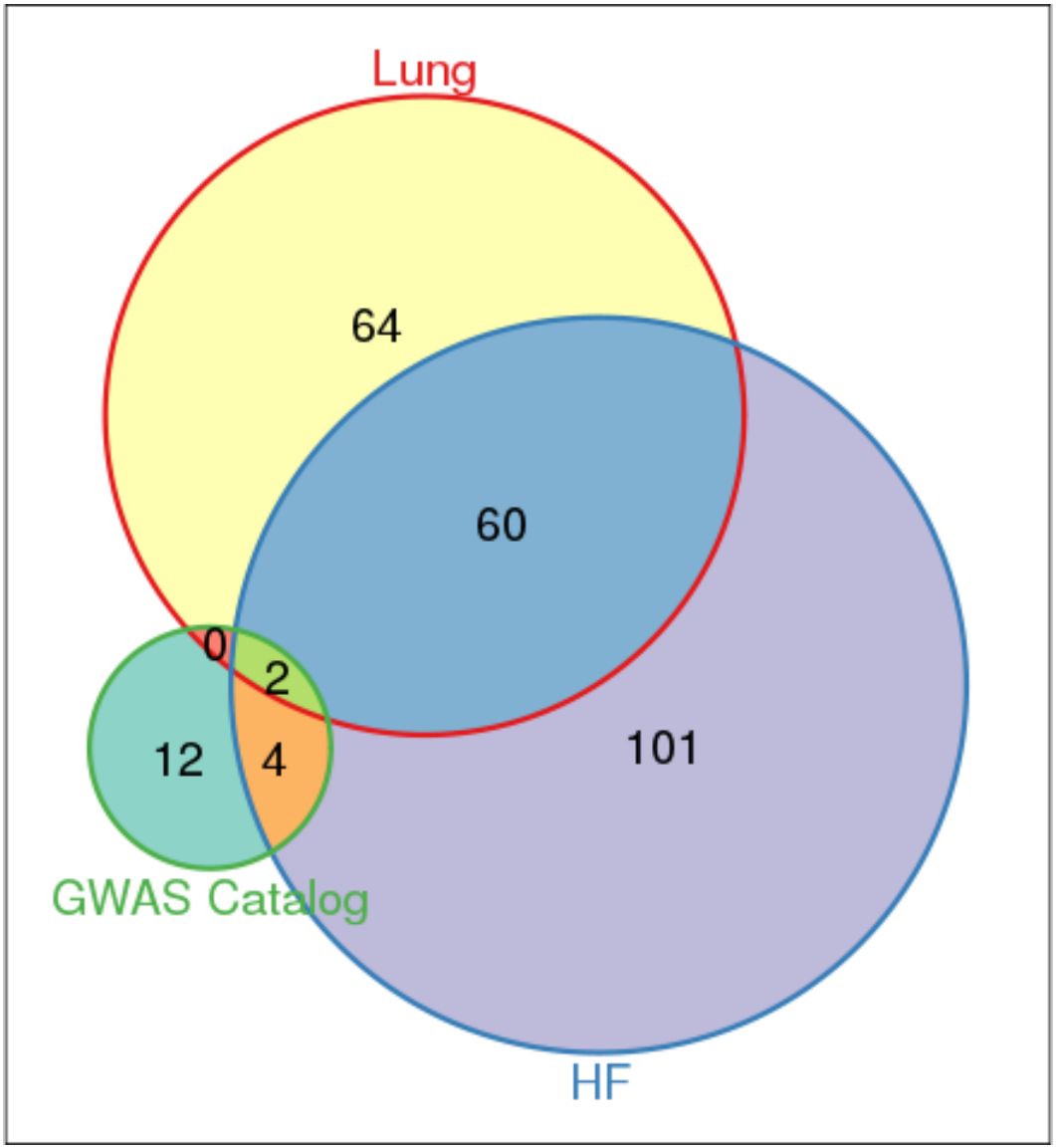
# From variants to gene

## Sets intersection



# From variants to gene

## Sets intersection



# From genes to pathways

## HF

| Pathway.identifier | Pathway.name | Entities.pValue | Submitted.entities.found |
|---|---|---|---|
| R-HSA-6809371 | Formation of the cornified envelope | 8.664749734155208E-6 | KRT71;KRT82;KRT80;KRT4;KRT3;KRT2;KRT1;KRT79;KRT8;KRT78;KRT7;KRT77;KRT76;KRT75;KRT86;KRT74;KRT85;KRT84;KRT83;KRT18;PKP2 |
| R-HSA-420499 | Class C/3 (Metabotropic glutamate/pheromone receptors) | 0.004953950850458244 | TAS2R7;TAS2R42 |
| R-HSA-6805567 | Keratinization | 0.014321908213508316 | KRT71;KRT82;KRT80;KRT4;KRT3;KRT2;KRT1;KRT79;KRT8;KRT78;KRT7;KRT77;KRT76;KRT75;KRT86;KRT74;KRT85;KRT84;KRT83;KRT18;PKP2 |
| R-HSA-72731 | Recycling of eIF2:GDP | 0.03954795734154837 | SLC5A8;EIF2B1;EIF2S3L |
| R-HSA-3249367 | STAT6-mediated induction of chemokines | 0.054077331642961024 | TBK1;STAT6 |
| R-HSA-6802948 | Signaling by high-kinase activity BRAF mutants | 0.05887835859772694 | RAP1B;VWF;PEBP1;KRAS;KSR2 |
| R-HSA-5625900 | RHO GTPases activate CIT | 0.06381045902536409 | CDKN1B;MYL6B;CIT |
| R-HSA-6802946 | Signaling by moderate kinase activity BRAF mutants | 0.06441268586272786 | RAP1B;VWF;PEBP1;KRAS;KSR2 |
| R-HSA-5674135 | MAP2K and MAPK activation | 0.06441268586272786 | RAP1B;VWF;PEBP1;KRAS;KSR2 |
| R-HSA-6802949 | Signaling by RAS mutants | 0.0660408405716818 | RAP1B;VWF;PEBP1;RASAL1;KRAS;KSR2 |
| R-HSA-6802955 | Paradoxical activation of RAF signaling by kinase inactive BRAF | 0.07025180102062634 | RAP1B;VWF;PEBP1;KRAS;KSR2 |
| R-HSA-139853 | Elevation of cytosolic Ca2+ levels | 0.08356236576501674 | P2RX7;P2RX4;P2RX2;ITPR2 |
| R-HSA-5250989 | Toxicity of botulinum toxin type G (BoNT/G) | 0.09637150792363747 | SYT1;VAMP1 |
| R-HSA-69091 | Polymerase switching | 0.1180671078261919 | RFC5;PRIM1 |
| R-HSA-69109 | Leading Strand Synthesis | 0.1180671078261919 | RFC5;PRIM1 |
| R-HSA-2995410 | Nuclear Envelope Reassembly | 0.1180671078261919 | ANKLE2;LEMD3;TMPO |
| R-HSA-2995383 | Initiation of Nuclear Envelope Reformation | 0.1180671078261919 | ANKLE2;LEMD3;TMPO |
| R-HSA-5635851 | GLI proteins bind promoters of Hh responsive genes to promote transcription | 0.12008284192932228 | GLI1 |
| R-HSA-8849470 | PTK6 Regulates Cell Cycle | 0.12008284192932228 | CDKN1B;CDK4 |
| R-HSA-110328 | Recognition and association of DNA glycosylase with site containing an affected pyrimidine | 0.12008284192932228 | SMUG1;TDG |

*Taula 5: Pathways*