Bioinformatics 2019-2020

Date: 15 January 2020

Project report

|  |
| --- |
| Bioinformatics@Data Science A.Y. 2019-2020  **Network Medicine project**  Simone Faricelli1, Lorenzo Germano1  1Group no. 09 Abstract Max 150 words. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. |

Basic introduction about the disease/process

Diabetes mellitus: More commonly referred to as "diabetes", a chronic disease associated with abnormally high levels of the sugar glucose in the blood.

Diabetes is due to one of two mechanisms: Inadequate production of insulin (which is made by the pancreas and lowers blood glucose), or Inadequate sensitivity of cells to the action of insulin.

The two main types of diabetes correspond to these two mechanisms and are called insulin dependent (type 1) and non-insulin dependent (type 2) diabetes. In type 1 diabetes there is no insulin or not enough of it. In type 2 diabetes, there is generally enough insulin but the cells upon which it should act are not normally sensitive to its action.

The signs and symptoms of both types of diabetes include increased urine output and decreased appetite as well as fatigue. Diabetes is diagnosed by blood glucose testing, the glucose tolerance test, and testing of the level of glycosylated hemoglobin (glycohemoglobin or hemoglobin A1C). The mode of treatment depends on the type of the diabetes.

The major complications of diabetes include dangerously elevated blood sugar, abnormally low blood sugar due to diabetes medications, and disease of the blood vessels which can damage the eyes, kidneys, nerves, and heart.

Seed genes

We started from the DisGeNet dataset, looking for every gene involved in the disease of Diabetes mellitus regarding the Homo Sapiens. We manually downloaded the whole dataset from the site and stored in the following table all the information about the above-mentioned genes using Matlab and selecting only the genes with the corresponding pathology code of Diabetes ('C0011853’).

|  |  |  |  |
| --- | --- | --- | --- |
| **Gene\_Symbol** | **Gene\_ID** | **Uniprot\_ID** | **Protein\_Name** |
| ACOX1 | 51 | Q15067 | acyl-CoA oxidase 1 |
| ADRA1A | 148 | P35348 | adrenoceptor alpha 1A |
| ADRB3 | 155 | P13945 | adrenoceptor beta 3 |
| AGT | 183 | P01019 | angiotensinogen |
| FAS | 355 | P25445 | Fas cell surface death receptor |
| STS | 412 | P08842 | steroid sulfatase |
| ATF3 | 467 | P18847 | activating transcription factor 3 |
| ATP2A2 | 488 | P16615 | ATPase sarcoplasmic/endoplasmic reticulum Ca2+ transporting 2 |
| ATP2A3 | 489 | Q93084 | ATPase sarcoplasmic/endoplasmic reticulum Ca2+ transporting 3 |
| BAX | 581 | Q07812 | BCL2 associated X, apoptosis regulator |
| BCL2 | 596 | P10415 | BCL2 apoptosis regulator |
| BCL2L1 | 598 | Q07817 | BCL2 like 1 |
| BDKRB1 | 623 | P46663 | bradykinin receptor B1 |
| CASP3 | 836 | P42574 | caspase 3 |
| CAT | 847 | P04040 | catalase |
| CAV1 | 857 | Q03135 | caveolin 1 |
| CAV3 | 859 | P56539 | caveolin 3 |
| CD68 | 968 | P34810 | CD68 molecule |
| CHRM2 | 1129 | P08172 | cholinergic receptor muscarinic 2 |
| CPT1A | 1374 | P50416 | carnitine palmitoyltransferase 1A |
| CPT1B | 1375 | Q92523 | carnitine palmitoyltransferase 1B |
| CYBA | 1535 | P13498 | cytochrome b-245 alpha chain |
| CYBB | 1536 | P04839 | cytochrome b-245 beta chain |
| CYP1A1 | 1543 | P04798 | cytochrome P450 family 1 subfamily A member 1 |
| ACE | 1636 | P12821 | angiotensin I converting enzyme |
| NQO1 | 1728 | P15559 | NAD(P)H quinone dehydrogenase 1 |
| EDN1 | 1906 | P05305 | endothelin 1 |
| ESRRA | 2101 | P11474 | estrogen related receptor alpha |
| ACSL1 | 2180 | P33121 | acyl-CoA synthetase long chain family member 1 |
| FOXO3 | 2309 | O43524 | forkhead box O3 |
| GCK | 2645 | P35557 | glucokinase |
| GPD2 | 2820 | P43304 | glycerol-3-phosphate dehydrogenase 2 |
| GPX1 | 2876 | P07203 | glutathione peroxidase 1 |
| GSR | 2936 | P00390 | glutathione-disulfide reductase |
| HK1 | 3098 | P19367 | hexokinase 1 |
| HMOX1 | 3162 | P09601 | heme oxygenase 1 |
| HSD11B1 | 3290 | P28845 | hydroxysteroid 11-beta dehydrogenase 1 |
| IAPP | 3375 | P10997 | islet amyloid polypeptide |
| ICAM1 | 3383 | P05362 | intercellular adhesion molecule 1 |
| ID1 | 3397 | P41134 | inhibitor of DNA binding 1, HLH protein |
| IFNG | 3458 | P01579 | interferon gamma |
| IGF1 | 3479 | P05019 | insulin like growth factor 1 |
| IL1B | 3553 | P01584 | interleukin 1 beta |
| IL6 | 3569 | P05231 | interleukin 6 |
| INSR | 3643 | P06213 | insulin receptor |
| PDX1 | 3651 | P52945 | pancreatic and duodenal homeobox 1 |
| IRS1 | 3667 | P35568 | insulin receptor substrate 1 |
| KCNJ11 | 3767 | Q14654 | potassium inwardly rectifying channel subfamily J member 11 |
| LEP | 3952 | P41159 | leptin |
| LEPR | 3953 | P48357 | leptin receptor |
| MAP3K5 | 4217 | Q99683 | mitogen-activated protein kinase kinase kinase 5 |
| MFGE8 | 4240 | Q08431 | milk fat globule EGF and factor V/VIII domain containing |
| MMP2 | 4313 | P08253 | matrix metallopeptidase 2 |
| MMP9 | 4318 | P14780 | matrix metallopeptidase 9 |
| MPO | 4353 | P05164 | myeloperoxidase |
| COX2 | 4513 | null | null |
| ND1 | 4535 | null | null |
| NEUROD1 | 4760 | Q13562 | neuronal differentiation 1 |
| NKX6-1 | 4825 | P78426 | NK6 homeobox 1 |
| NOS2 | 4843 | P35228 | nitric oxide synthase 2 |
| NOS3 | 4846 | P29474 | nitric oxide synthase 3 |
| SERPINE1 | 5054 | P05121 | serpin family E member 1 |
| PAX6 | 5080 | P26367 | paired box 6 |
| PCK1 | 5105 | P35558 | phosphoenolpyruvate carboxykinase 1 |
| PCSK2 | 5126 | P16519 | proprotein convertase subtilisin/kexin type 2 |
| PDK4 | 5166 | Q16654 | pyruvate dehydrogenase kinase 4 |
| PFKM | 5213 | P08237 | phosphofructokinase, muscle |
| PKLR | 5313 | P30613 | pyruvate kinase L/R |
| PPARA | 5465 | Q07869 | peroxisome proliferator activated receptor alpha |
| PPARG | 5468 | P37231 | peroxisome proliferator activated receptor gamma |
| PRKCA | 5578 | P17252 | protein kinase C alpha |
| PRKCD | 5580 | Q05655 | protein kinase C delta |
| PRKCE | 5581 | Q02156 | protein kinase C epsilon |
| PTGS2 | 5743 | P35354 | prostaglandin-endoperoxide synthase 2 |
| RELA | 5970 | Q04206 | RELA proto-oncogene, NF-kB subunit |
| REN | 5972 | P00797 | renin |
| S100A6 | 6277 | P06703 | S100 calcium binding protein A6 |
| CCL20 | 6364 | P78556 | C-C motif chemokine ligand 20 |
| CX3CL1 | 6376 | P78423 | C-X3-C motif chemokine ligand 1 |
| SLC2A2 | 6514 | P11168 | solute carrier family 2 member 2 |
| SLC2A4 | 6517 | P14672 | solute carrier family 2 member 4 |
| SLC9A1 | 6548 | P19634 | solute carrier family 9 member A1 |
| SLC9A3 | 6550 | P48764 | solute carrier family 9 member A3 |
| SNAP25 | 6616 | P60880 | synaptosome associated protein 25 |
| SOD1 | 6647 | P00441 | superoxide dismutase 1 |
| SOD2 | 6648 | P04179 | superoxide dismutase 2 |
| SREBF1 | 6720 | P36956 | sterol regulatory element binding transcription factor 1 |
| TGFB1 | 7040 | P01137 | transforming growth factor beta 1 |
| TIMP1 | 7076 | P01033 | TIMP metallopeptidase inhibitor 1 |
| TIMP2 | 7077 | P16035 | TIMP metallopeptidase inhibitor 2 |
| TNF | 7124 | P01375 | tumor necrosis factor |
| TNFRSF1A | 7132 | P19438 | TNF receptor superfamily member 1A |
| TP53 | 7157 | P04637 | tumor protein p53 |
| UCP2 | 7351 | P55851 | uncoupling protein 2 |
| VEGFA | 7422 | P15692 | vascular endothelial growth factor A |
| YWHAH | 7533 | Q04917 | tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein eta |
| AOC3 | 8639 | Q16853 | amine oxidase copper containing 3 |
| IRS2 | 8660 | Q9Y4H2 | insulin receptor substrate 2 |
| S1PR4 | 8698 | O95977 | sphingosine-1-phosphate receptor 4 |
| AIFM1 | 9131 | O95831 | apoptosis inducing factor mitochondria associated 1 |
| S1PR2 | 9294 | O95136 | sphingosine-1-phosphate receptor 2 |
| PPARGC1A | 10891 | Q9UBK2 | PPARG coactivator 1 alpha |
| SIRT1 | 23411 | Q96EB6 | sirtuin 1 |
| FGF21 | 26291 | Q9NSA1 | fibroblast growth factor 21 |
| S1PR5 | 53637 | Q9H228 | sphingosine-1-phosphate receptor 5 |
| GPAM | 57678 | Q9HCL2 | glycerol-3-phosphate acyltransferase, mitochondrial |
| ACOT1 | 641371 | Q86TX2 | acyl-CoA thioesterase 1 |
| NCF1 | 653361 | P14598 | neutrophil cytosolic factor 1 |

Summary on interaction data

Explain briefly the methods you followed to get the information about the interaction data and add the related table. Refer clearly to different files (i.e. when necessary.

Sample numbered list, if necessary.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

1. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

* Sample bullet list, if necessary.
* The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.
* The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

**Table 1.**Sample table. This should be the table format, add/remove columns and rows according to the data to be shown.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |S| | Predicted cost | Timing | Predicted speed | Speed |
| 1 | S219.20(100%) | 68m43s | 1.00 | 1.00 |
| 2 | 29.10+219.10(~50%) | 35m13s | 2.00 | 1.95 |
| 4 | 219.20(100%) | 68m43s | 1.00 | 1.00 |
| 10 | 29.10+219.10(~50%) | 35m13s | 2.00 | 1.95 |
| 20 | 219.20(100%) | 68m43s | 1.00 | 9.5 |

This is table foot note sample text This is table foot note sample text This is table foot note sample text

# Interactomes data

Explain briefly the methods you followed to build the intersection interactome and add the related tables/charts.

# Enrichment analysis

Explain briefly the methods you followed to carry out the enrichment analysis and add the related tables/charts.

# Notes and comments

References (if any, this is the format to be used)

Alexandrescu,A. (2001) Modern C++ Design: Generic Programming and Design Patterens Applied. Addision Wesley Professional, Boston.

Dormand,J.R. and Prince,P.J. (1980) A family of embedded Runge–Kutta formulae. *J. Comp. Appl. Math.*, **6**, 19–26.