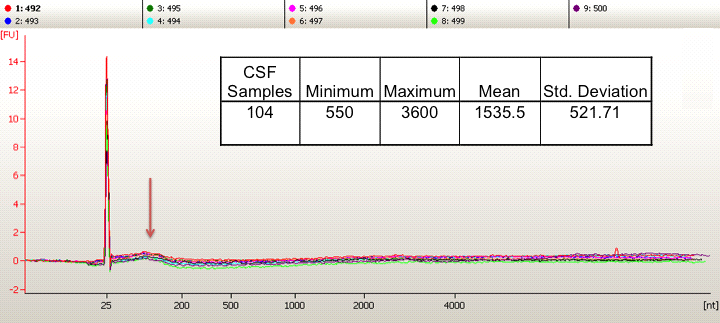
**Supplementary Table 1.** *UWADRC CSF Data Dictionary.* Basic dataset information and study inclusion criteria for test subjects.

|  |  |  |  |
| --- | --- | --- | --- |
| **Category information** | **Variable name** | **Definition** | **Code and data range** |
| **IDs** | **EntityID** | Unique ID for each Subject assigned by database manager | 0-99999 |
|  | **SampleID (link key)** | Subject ID # for ApoE/CSF or BioMarQ tudy on the specimen tube and pull sheet | 0-99999 |
|  | **Lpdate (link key)** | Date of LP | MM/DD/YYYY |
|  | **ScreenDate** | Date of clinical visit was conducted. | MM/DD/YYYY |
| **Demographics** | **Byr** | Subject Birth Year | 1900-2022 |
|  | **AgeAtEntry** | Age of subject at entry into AC/BQ study | 18-99 |
|  | **Age of Scrdate** | Age of clinical visit | 18-99 |
|  | **OnsetAge** | Age at onset (first sx of impairment) | 18-98 (99=n/a) |
|  | **Gender** | Sex of subject | 1, 2 (1=male, 2=female) |
|  | **RACE** | Race of subject | 1=white, 2=black,3=American Indian/Alaskan Native, 4=Asian/Pacific Islander, 5=Other, 9=Missing/unknown |
|  | **HISPANIC** | Is subject Spanish/Latino/Hispanic | 1=Yes, 2=No, 9=Unknown |
|  | **Education** | Years of formal education | 00-99 |
| **Diagnosis** | **Group, PrimaryDx** | Subject's Group/Primary Diagnosis | 1=young control (18-39),2=middle-aged control (40-64),3=older control (65+),4=AD,5=MCI, 6=Other |
| **APOE genotype** | **APOE** | APOE genotype | x/x |
| **Vital sign** | **BMI13** | Body mass index | 18-46 (BMI outside of 18-32 waived for eligibility) |
| **CDR** | **CDRS16** | Overall CDR score | 0=no dementia,0.5=uncertain,1=mild,2=moderate ,3=severe, 4=very severe ,5=terminal dementia |
| **MMSE** | **MMSWORLD** | MMSE Total score (with "world") | 0-30 |
| **Logic Memory** | **DELSCORE** | Delayed recall score | 0-25, 96=not given due to cognitive impairment, 99=not administered due to other reasons |
| **CSF measures** | **BQCSFRBCS** | CSF RBCs (clinical lab results ) | Clinical Ref. Range <6 |
|  | **BQCSFPROTEIN** | CSF Total Protein (clinical lab results ) | Clinical Ref. Range 15-45 |
|  | **BQCSFGLUCOSE** | CSF Glucose (clinical lab results ) | Clinical Ref. Range 40-70 |
|  | **CSF draw #** | Gradient of CSF | 1-50 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Supplementary Table 1. UWADRC CSF data dictionary – Expanded dataset** | | | |
| **Category information** | **Variable name** | **Definition** | **Code and data range** |
| **IDs** | **EntityID** | Unique ID for each Subject assigned by database manager | 0-99999 |
|  | **SampleID (link key)** | Subject ID # for ApoE/CSF (AC), BioMarQ study (BQ), or other studies on the specimen tube and pull sheet | 0-99999 |
|  | **Lpdate (link key)** | Date of LP | MM/DD/YYYY |
|  | **ScreenDate** | Date of clinical visit was conducted. | MM/DD/YYYY |
| **Demographics** | **Byr** | Subject Birth Year | 1900-2022 |
|  | **AgeAtEntry** | Age of subject at entry into AC/BQ study | 18-99 |
|  | **Age of Scrdate** | Age of clinical visit | 18-99 |
|  | **OnsetAge** | Age at onset (first sx of impairment) | 18-98 (99=n/a) |
|  | **AGEDX** | Age at dementia diagnosis | 19-98 (99=n/a) |
|  | **Gender** | Sex of subject | 1, 2 (1=male, 2=female) |
|  | **RACE** | Race of subject | 1=white, 2=black,3=American Indian/Alaskan Native, 4=Asian/Pacific Islander, 5=Other,9=Missing/unknown |
|  | **HISPANIC** | Is subject Spanish/Latino/Hispanic | 1=Yes, 2=No, 9=Unknown |
|  | **Marital** | Current Marital status. | 1=married,2=widowed,3=divorced,4=separated,5=never married,6=other,9=missing/unknown |
|  | **Education** | Years of formal education | 00-99 |
| **Diagnosis** | **Group, PrimaryDx** | Subject's Group/Primary Diagnosis | 1=young control (18-39), 2=middle-aged control (40-64), 3=older control (65+), 4=AD, 5=MCI, 6=Other |
|  | **SITE** | Location where subject tested | blank=UW Site, 5=UC Davis, 6=UCSD, 7=Indiana U, 8=OHSU, 9=Upenn |
| **APOE genotype** | **APOE** | APOE genotype | x/x |
| **Family history** | **MEMPROB** | Family history of memory problems | 1=Yes, 2=No, 9=Unknown |
|  | **AFFECT** | Three or more affected blood relatives, living or deceased | 1=Yes, 2=No, 9=Unknown |
| **ETOH/Smoking history** | **SUBSTANC** | drink > 2 alcoholic drinks/d or 10/wk, or use of recreation drug | 1=Yes, 2=No |
|  | **CURRTOB** | Current tobacco use | 1=Yes, 2=No |
|  | **PREVTOB** | Past tobacco use | 1=Yes, 2=No |
| **Cognitive-enhancing agents** | **COGAGENT** | Subject ever taken any cognitive-enhancing agents? | 1=Yes, 2=No |
| **PD related** | **NO\_PD** | No parkinsonian features | 1=Yes, 2=No |
|  | **UPDRSTOT** | UPDRS total score | 0-108 |
| **Vital signs** | **BP** | Blood pressure (systolic) | 0-999 |
|  | **BPU** | Blood pressure (diastolic) | 0-999 |
|  | **PULSE** | Heart beats per minute | 0-999 |
|  | **WEIGHT** | Weight in lbs | 0-999.9 |
|  | **BMI13** | Body mass index (BMI) | 18-46 (outside of 18-32 waived for eligibility) |
| **CDR** | **CDRS16** | Overall CDR score | 0=no dementia,0.5=uncertain,1=mild,2=moderate ,3=severe, 4=very severe ,5=terminal dementia |
| **Function measurements** | **BQTOTALFAQ** | BQ ONLY FUNCTIONAL ACTIVITIES QUESTIONNAIRE (FAQ) Total | 0-30 |
|  | **BQFAQNS** | BQ ONLY Number of "N" responses on FAQ. | 0-10 |
| **NPI** | **CARETOTAL** | Total score of caregiver distress | 0-144 |
|  | **NPITOTAL** | Total score of NPI (frequency x severity) | 0-144 |
| **Depression** | **GDSS34** | Geriatric Depression Scale total score | 0-15 (99=unable to complete) |
|  | **HAMTOT** | AC ONLY Hamilton Depression Rating Scale total score | 0-98 (99=unable to complete) |
|  | **BDITOTAL** | AC ONLY Beck Depression Inventory total score | 0-63 |
| **MMSE** | **MMSWORLD** | MMSE Total score (with "world") | 0-30 |
| **Logic Memory** | **IMMSCORE** | Logical Memory 1 Immediate recall score | 0-25, 96=not given due to cognitive impairment, 99=not administered due to other reasons |
|  | **DELSCORE** | Delayed recall score | 0-25, 96=not given due to cognitive impairment, 99=not administered due to other reasons |
| **Digit Span** | **BQFDIGCORR** | BQ ONLY Digit Span Forward Total Correct | 0-14 |
|  | **BQFDIGLENGTH** | BQ ONLY Digit Span Forward Length | 0-9 |
|  | **BQBDIGCORR** | BQ ONLY Digit Span Backward Total Correct | 0-14 |
|  | **BQBDIGLENGTH** | BQ ONLY Digit Span Backward Length | 0-9 |
| **Verbal fluency** | **ANIMAL** | Category Fluency (animals) total score | 0-999 |
|  | **LETTER** | AC ONLY Letter Fluency (letter "S") total score | 0-999 |
|  | **BQVEGETABLE** | BQ ONLY Category Fluency (vegetables) total score. | 0-999 |
| **Trails Making Test** | **TMTTIMEA** | Trail Making Test (Part A) time (seconds) | 0-300 - Over 150s considered a fail. (995 = discont. due to physical problem, 996 = discontinued/not attempted due to cognitive impairment, 997 = other problem, 998 = refused) |
|  | **TMTTIMEB** | Trail Making Test (Part B) time (seconds) | 0-300 - Over 300s considered a fail. (995 = discont. due to physical problem, 996 = discontinued/not attempted due to cognitive impairment, 997 = other problem, 998 = refused) |
| **Digit Symbol test** | **BQDIGSYMTOTAL** | BQ ONLY Digit Symbol Total Score (Number completed - Incorrect) | 0-93 (96 = not completed/attempted due to cognitive impairment, 98 = verbal refusal) |
| **Boston Naming** | **BNTOTALCORR** | BQ ONLY Total Correct Responses  A + C   (maximum of 30) | 0-99 |
|  | **BNVersion** | BQ ONLY, Boston Naming Test Used. | Null = Correct Version, 1=Version 1, 2=Other.  Correct = 1-59 odd, Version 1 = 30-60 (skip noose), 2 = any Other combination of items given |
| **CSF measures** | **BQCSFRBCS** | CSF RBCs (clinical lab results ) | Clinical Ref. Range <6 |
|  | **BQCSFPROTEIN** | CSF Total Protein (clinical lab results ) | Clinical Ref. Range 15-45 |
|  | **BQCSFGLUCOSE** | CSF Glucose (clinical lab results ) | Clinical Ref. Range 40-70 |
|  | **CSF draw #** | Gradient of CSF | 1-50 |

**Supplementary Figure 1.** *RNA content of CSF.* The CSF RNA profile was visualized using Agilent’s RNA 6000 Pico Kit. The table shows the amount of RNA present in the samples, with minimum and maximum RNA amounts ranging from 550 to 3600 picograms with a mean of 1535.5 picograms in all samples. The RNA fragment size in nucleotides (log scale) was plotted against fluorescence units (FU) to measure the RNA concentration.



**Supplementary Table 2.** *CSF extracellular RNA profiling.* Using our optimized method and directional RNA library preparation kit, we prepared libraries and sequenced them on the Illumina HiSeq 2000 platform. The average of trimmed reads was 81 million per sample (96 %), of which 82.7 % aligned to the human reference genome (UCSC\_hg19).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sample Name** | **# of reads** | **# of trimmed reads** | **Trimmed reads (%)** | **Alignment STAR (%)** |
| B1 | 83712781 | 82640236 | 98.70 | 86.07 |
| A1 | 84253862 | 83242682 | 98.80 | 86.96 |
| A2 | 81349142 | 80159194 | 98.55 | 82.44 |
| A3 | 67626000 | 65392108 | 96.65 | 82.41 |
| A4 | 71229859 | 68604214 | 96.30 | 82.47 |
| A5 | 66514069 | 62879623 | 94.50 | 81.88 |
| B2 | 52695677 | 49213572 | 93.40 | 81.65 |
| A6 | 58709907 | 56293791 | 95.90 | 82.70 |
| A7 | 71906039 | 69387223 | 96.50 | 83.53 |
| B3 | 63464686 | 61044737 | 96.15 | 82.73 |
| B4 | 75056867 | 72041531 | 96.00 | 83.33 |
| A8 | 60769421 | 58998451 | 97.05 | 83.22 |
| B5 | 64397539 | 61770448 | 95.90 | 82.39 |
| B6 | 65757953 | 63399372 | 96.40 | 82.68 |
| A9 | 56340893 | 54799270 | 97.25 | 83.63 |
| B7 | 64925540 | 62783336 | 96.70 | 82.69 |
| A10 | 66695977 | 64398236 | 96.55 | 82.73 |
| A11 | 76892781 | 74920902 | 97.45 | 82.51 |
| A12 | 58617176 | 56474861 | 96.35 | 82.29 |
| A13 | 54009544 | 51708939 | 95.75 | 82.98 |
| B8 | 71294764 | 69122634 | 96.95 | 82.84 |
| B9 | 61816117 | 59795173 | 96.85 | 82.53 |
| B10 | 97581493 | 90715379 | 92.25 | 81.61 |
| B11 | 87460057 | 84748795 | 96.90 | 87.60 |
| B12 | 63689824 | 58435746 | 91.75 | 87.88 |
| B13 | 134519335 | 133108700 | 98.95 | 84.51 |
| B14 | 144765645 | 143246896 | 98.95 | 86.73 |
| A14 | 82374351 | 79373038 | 96.35 | 82.69 |
| A15 | 64000497 | 62948901 | 98.30 | 82.98 |
| A16 | 89176574 | 85742748 | 96.30 | 81.24 |
| B15 | 105379688 | 101563280 | 96.25 | 80.97 |
| B16 | 99421053 | 96924385 | 97.50 | 83.17 |
| B17 | 95829619 | 93011822 | 97.05 | 82.29 |
| B18 | 105259505 | 101314510 | 96.25 | 82.29 |
| A17 | 86373546 | 83008636 | 96.10 | 82.55 |
| B19 | 90396184 | 86453400 | 95.65 | 83.58 |
| B20 | 100216026 | 97228366 | 97.20 | 82.10 |
| B21 | 68432385 | 66154996 | 96.85 | 81.80 |
| A18 | 69888810 | 67408222 | 96.35 | 82.03 |
| A19 | 51814633 | 50445447 | 97.35 | 82.05 |
| A20 | 76167172 | 73617115 | 96.65 | 80.35 |
| A21 | 85425116 | 81412622 | 95.30 | 80.20 |
| B22 | 60762092 | 58484669 | 96.25 | 79.91 |
| A22 | 66956433 | 64714451 | 96.65 | 79.54 |
| B23 | 74756000 | 71991844 | 96.30 | 79.69 |
| B24 | 58002799 | 56060783 | 96.65 | 80.55 |
| A23 | 76766392 | 64866171 | 84.50 | 82.46 |
| B25 | 80319781 | 77068676 | 95.95 | 81.40 |
| A24 | 113323040 | 108109987 | 95.40 | 83.61 |
| A25 | 114270484 | 109128151 | 95.50 | 82.62 |
| A26 | 114845325 | 110998027 | 96.65 | 81.28 |
| B26 | 113749281 | 110279421 | 96.95 | 82.02 |
| B27 | 140318183 | 133722702 | 95.30 | 83.15 |
| A27 | 141441600 | 138117745 | 97.65 | 81.81 |
| A28 | 130542517 | 121535083 | 93.10 | 82.58 |
| A29 | 114220131 | 105710612 | 92.55 | 82.68 |
| A30 | 134354388 | 132137225 | 98.35 | 87.45 |
| Average | 84400641 | 81208054 | 96.15 | 82.74 |

**Supplementary Figure 2.** *cDNA synthesis of small fragmented RNA.* CSF RNAs were polyA-tailed using polyA polymerase and cDNA was synthesized using Oligo (dT) - adapter primers. These primers contain an oligo(dT) stretch and a specific sequence that is recognized by the universal PCR primers used in the PCR reaction.

|  |
| --- |
|  |

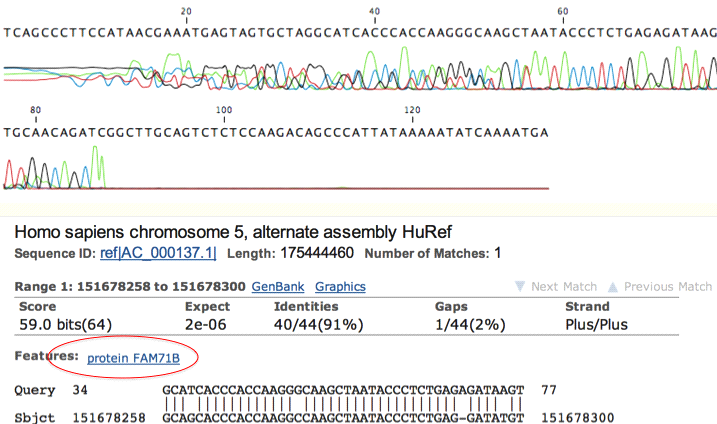
**Supplementary Figure 3.** *Modified real-time PCR to validate RNA-seq data.* Specific IGV-designed forward and a universal reverse primer were used to amplify gene targets. CSF extracellular RNA is small in size, so it was necessary to tail the RNA prior to cDNA synthesis.

|  |
| --- |
|  |

**Supplementary Table 3.** *Candidate genes for real-time PCR (qRT-PCR) validation and their forward primer sequences.*

|  |  |  |
| --- | --- | --- |
| **Target gene**  **(official name)** | **Transcript-specific primer** | **Gene Type** |
| *UBC* | GTGGTGCGTCCAGAGAGAC | protein\_coding |
| *SSTR5* | CGTCAACCAGTTCACCAGTG | protein\_coding |
| *RP5-1051H14.2* | CATGCTGAAGGAGATGGTGC | lincRNA |
| *RP3-467N11.1* | GGTGTCAAGGCCTGTTTTCC | antisense |
| *PTPRC* | GCTGAGGGTGGTTGTCAATG | protein\_coding |
| *PARK7* | AACCGGAAGGGCCTGATAG | protein\_coding |
| *NDUFV2* | AAGGTTGGGGAGACTACACC | protein\_coding |
| *LRRK2* | AGG CTG AAC AAT GTC CAG GA | protein\_coding |
| *LPAR1* | GCCAGTGAGAGTGTGGGT | protein\_coding |
| *EZH2* | ATGATGGAGACGATCCTGAA | protein\_coding |
| *DNMT1* | CTG CCA AAC GGA AAC CTC AG | protein\_coding |
| *CTD-2066L21.2* | CTCATGGCCTGCATCATTCC | lincRNA |
| *CCR3* | CAGCTGCTGTGGATTGGATT | protein\_coding |
| *BMP7* | TTTCCTTTCGCACAGACACC | protein\_coding |
| *AL356309.1* | AAGGCTAGGGTTGCCAGATG | miRNA |
| *AC010127.3* | TCTCCCTCCACACCTCCC | antisense |
| *MAPK11* | AGACAGGCTCCCACAGTG | protein\_coding |
| *uc001lva.4* | TGA AGC TAG AGG AAC CAG ACC | lincRNA |
| *AC079630.4* | CAC GGA AGG TCT GGG AAC T | lincRNA |
| *LCK* | GGTTACCTACGAAGGCTCCAA | protein\_coding |
| *NGF* | CAGGAAGGCTGTGAGAAGAG | protein\_coding |
| *SNCA* | TGGTGAGCGAAACAGAAGCC | protein\_coding |
| *SCN9A* | AGAGGGGTACACCTGTGTGAA | protein\_coding |

**Supplementary Figure 4.** *Sanger sequencing of Real-time PCR products.* Real-time PCR was performed by using specific primers targeting *FAM71b*. The qRT-PCR amplicon was confirmed with Sanger sequencing and was successfully blasted to the reference sequence in NCBI.



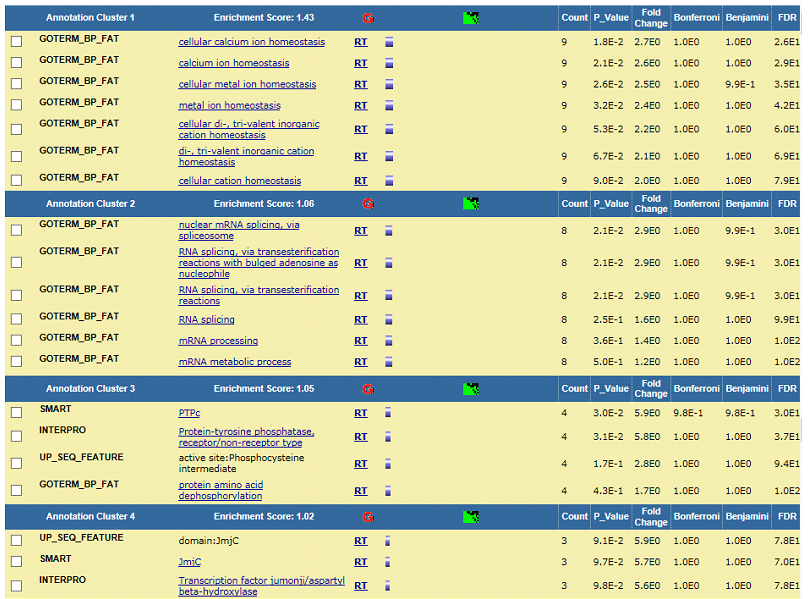
**Supplementary Table 4.** *Distribution of gene types present in the RNA-seq dataset.* After filtering, 3521 transcripts were detected using RNA-seq. Among them, 2862 (81.3 %) protein-coding and 320 (9.1 %) non-coding genes were identified.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **Frequency** | **Percent** | **Valid Percent** | **Cumulative Percent** |
| **Valid** | **antisense** | 127 | 3.6 | 3.6 | 3.6 |
| **lincRNA** | 182 | 5.2 | 5.2 | 8.8 |
| **miRNA** | 11 | .3 | .3 | 9.1 |
| **misc\_RNA** | 18 | .5 | .5 | 9.6 |
| **Mt\_rRNA** | 2 | .1 | .1 | 9.7 |
| **other** | 38 | 1.1 | 1.1 | 10.7 |
| **protein\_coding** | 2862 | 81.3 | 81.3 | 92.0 |
| **pseudogene** | 201 | 5.7 | 5.7 | 97.7 |
| **rRNA** | 36 | 1.0 | 1.0 | 98.8 |
| **sense\_intronic** | 7 | .2 | .2 | 98.9 |
| **sense\_overlapping** | 4 | .1 | .1 | 99.1 |
| **snoRNA** | 7 | .2 | .2 | 99.3 |
| **snRNA** | 26 | .7 | .7 | 100.0 |
| **Total** | 3521 | 100 | 100 |  |

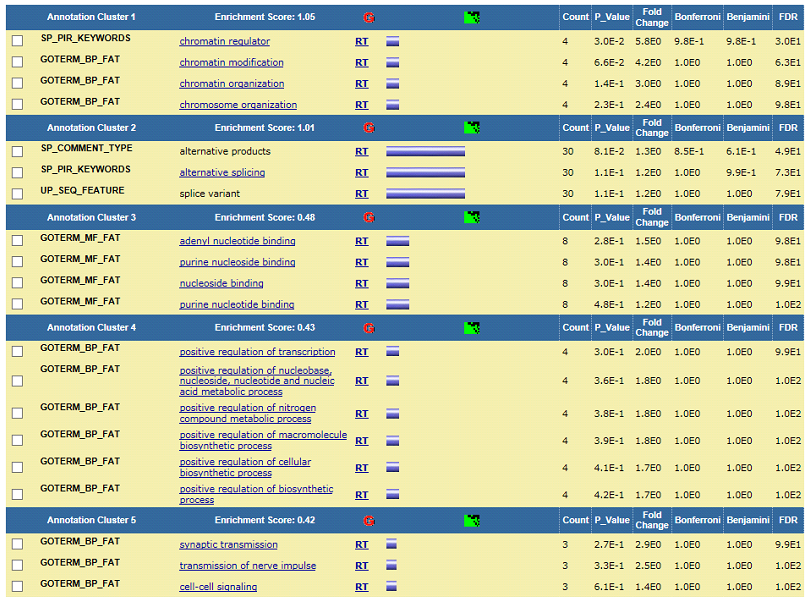
**Supplementary Table 5.** *Distribution of gene types in genes that are significantly differentially expressed between PD patients and healthy controls.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **Frequency** | **Percent** | **Valid Percent** | **Cumulative Percent** |
| **Valid** | **antisense** | 9 | 4.5 | 4.5 | 4.5 |
| **lincRNA** | 12 | 6.0 | 6.0 | 10.4 |
| **miRNA** | 6 | 3.0 | 3.0 | 13.4 |
| **misc\_RNA** | 4 | 2.0 | 2.0 | 15.4 |
| **other** | 2 | 1.0 | 1.0 | 16.4 |
| **protein\_coding** | 142 | 70.6 | 70.6 | 87.1 |
| **pseudogene** | 15 | 7.5 | 7.5 | 94.5 |
| **rRNA** | 5 | 2.5 | 2.5 | 97.0 |
| **sense\_intronic** | 1 | .5 | .5 | 97.5 |
| **snoRNA** | 2 | 1.0 | 1.0 | 98.5 |
| **snRNA** | 3 | 1.5 | 1.5 | 100.0 |
| **Total** | 201 | 100 | 100 |  |

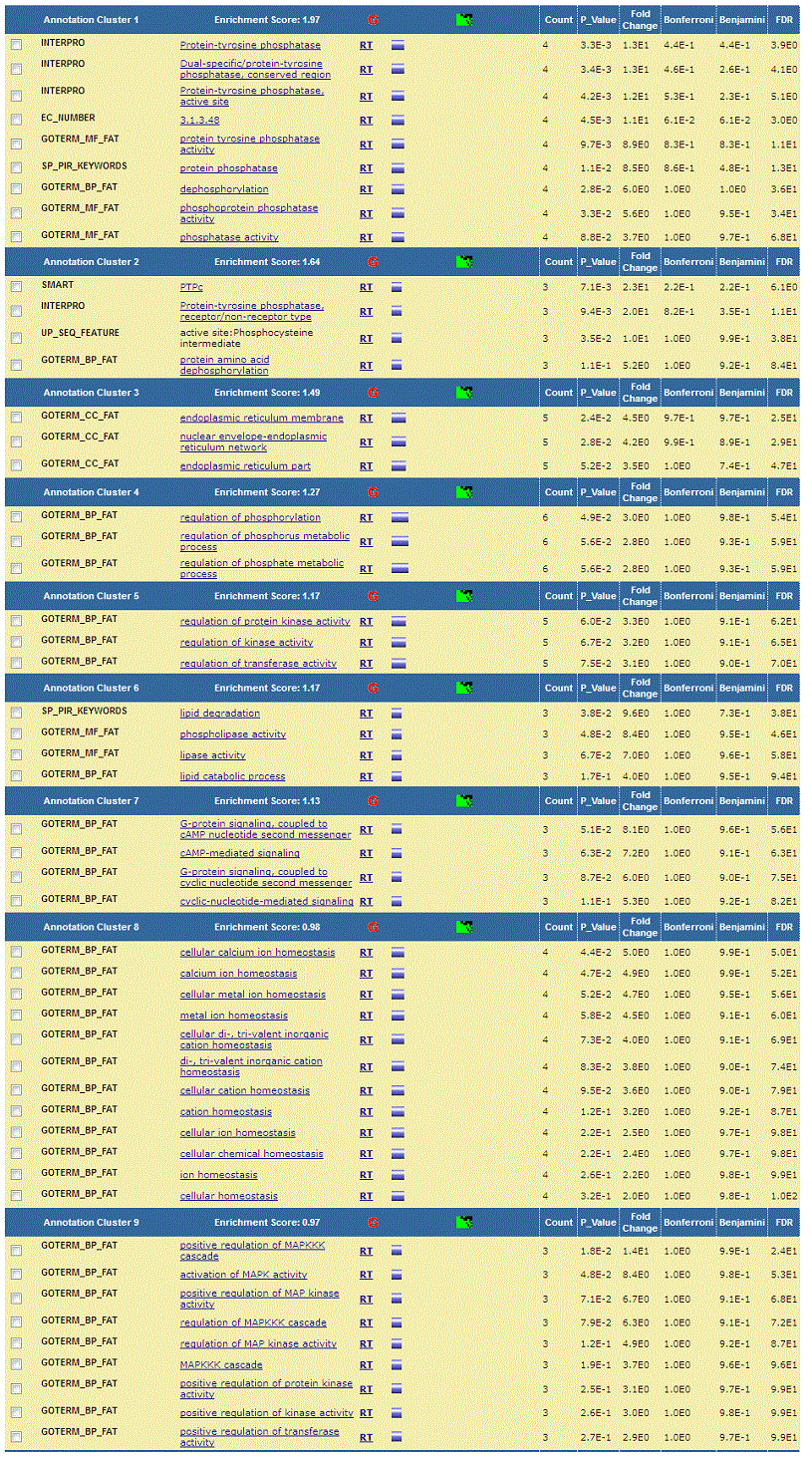
**Supplementary Table 6.** *Functional annotation clustering in 467 differentially expressed genes.* Differential expression analysis (p < 0.05) produced 467 DEGs. Functional annotation clustering using DAVID identified the molecular mechanisms affected in PD patients. DEGs were mainly enriched for genes involved in [cellular calcium ion homeostasis](http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0006874" \t "_blank), [nuclear mRNA splicing via the spliceosome](http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0000398" \t "_blank) and [protein-tyrosine phosphatases (receptor/ non-receptor type](http://www.ebi.ac.uk/interpro/entry/IPR000242" \t "_blank)s).

****

**Supplementary Table 7.** *Functional annotation clustering in differentially expressed genes.* Differential expression analysis using (|log FC| > 1, p < 0.05 and FDR < 0.1) identified 92 up-regulated DEGs in RNA-seq. Functional annotation clustering with DAVID identified molecular mechanisms affected by differential gene expression. Up-regulated genes were mostly enriched for [chromatin regulator](http://www.uniprot.org/keywords/?query=chromatin%20regulator" \t "_blank) genes.



**Supplementary** **Table 8.** *Functional annotation clustering in differentially expressed genes.* Differential expression analysis (|log FC| > 1, p < 0.05 and FDR < 0.1) identified 109 down-regulated DEGs in RNA-seq. Functional annotation clustering using DAVID identified molecular processes implicated in differential gene expression. Down-regulated genes were mainly enriched for genes involved in dephosphorylation, [protein-tyrosine phosphatase](http://www.ebi.ac.uk/interpro/entry/IPR000387" \t "_self)s, endoplasmic reticulum membrane, and regulation of phosphorylation and phospholipase activity pathways.



**Supplementary Figure 5.** *Protein-protein interaction (PPI) network.* A total of 201 DEGs were mapped to the STRING database; the hub protein was selected according to the node degree and screened significant interactions with a score > 0.7. By integrating these relationships, we constructed interaction networks amongst interacting proteins.

