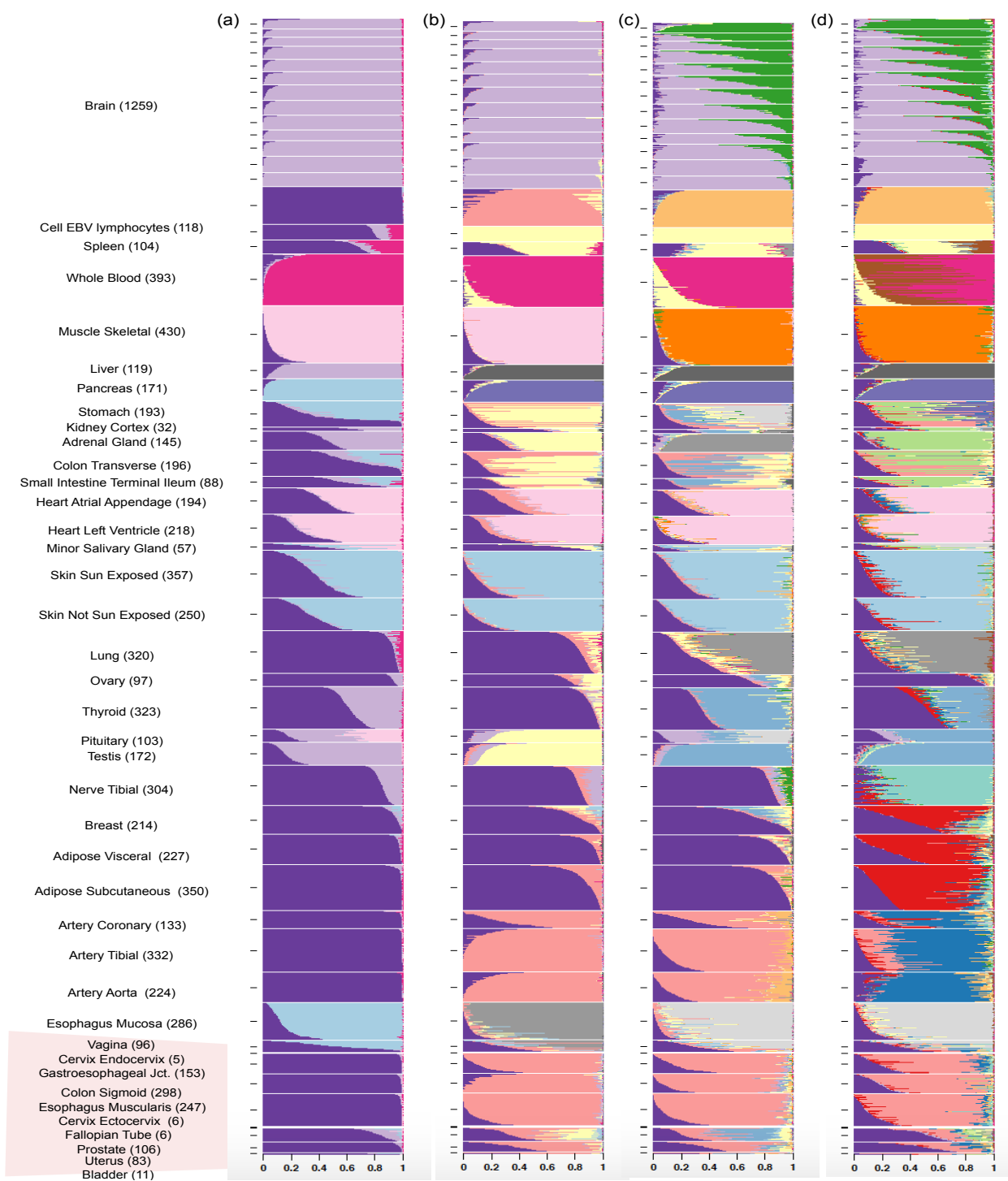
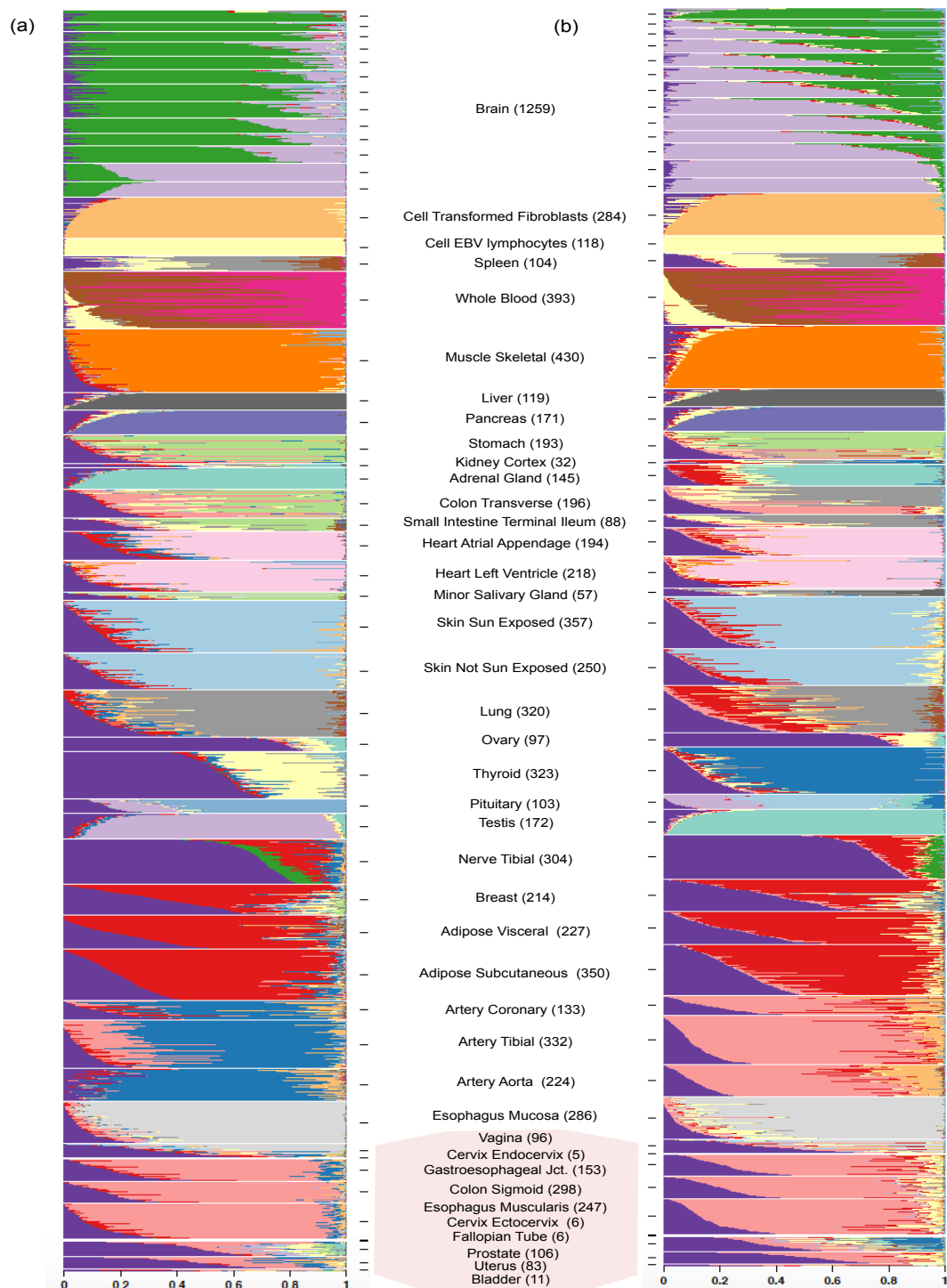


## 1 Supplementary figures

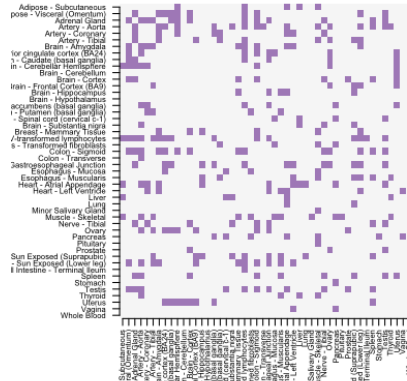
**S1 Fig. Structure plot of GTEx V6 all tissue samples for (a)  $K = 5$ , (b)  $K = 10$ , (c)  $K = 15$ , (d)  $K = 20$ .** Some tissues form a separate cluster from the rest of the tissues from  $K = 5$  onwards (for example: Whole Blood, Skin), whereas some tissues only form a distinctive subgroup only at  $K = 20$  (for example: Arteries).



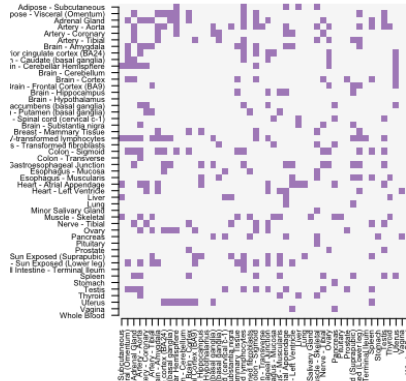
**S2 Fig. Structure plot of GTEx V6 all tissue samples K=20 in 2 runs under the thinning parameters settings (a)  $p_{thin} = 0.01$  and (b)  $p_{thin} = 0.0001$ .** The patterns in two plots closely correspond to the plot in Fig 1(a), though there are a few differences from the unthinned version.



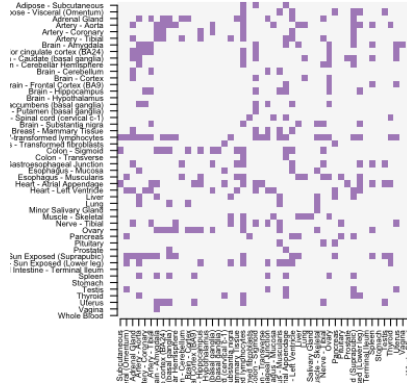
**S3 Fig. A comparison of “accuracy” of hierarchical vs. model-based clustering on thinned GTEx data, with thinning parameter  $p_{thin} = 0.01$  and  $p_{thin} = 0.0001$ . For each pair of tissues from the GTEx data we assessed whether or not each clustering method (with  $K = 2$  clusters) separated the samples according to their actual tissue of origin, with successful separation indicated by a filled square. Thinning deteriorates accuracy compared with the unthinned data (Fig 2), but even then the model-based method remains more successful than the hierarchical clustering in separating the samples by tissue or origin.**



(a) hierarchy thin 0.01



(b) GoM thin 0.01

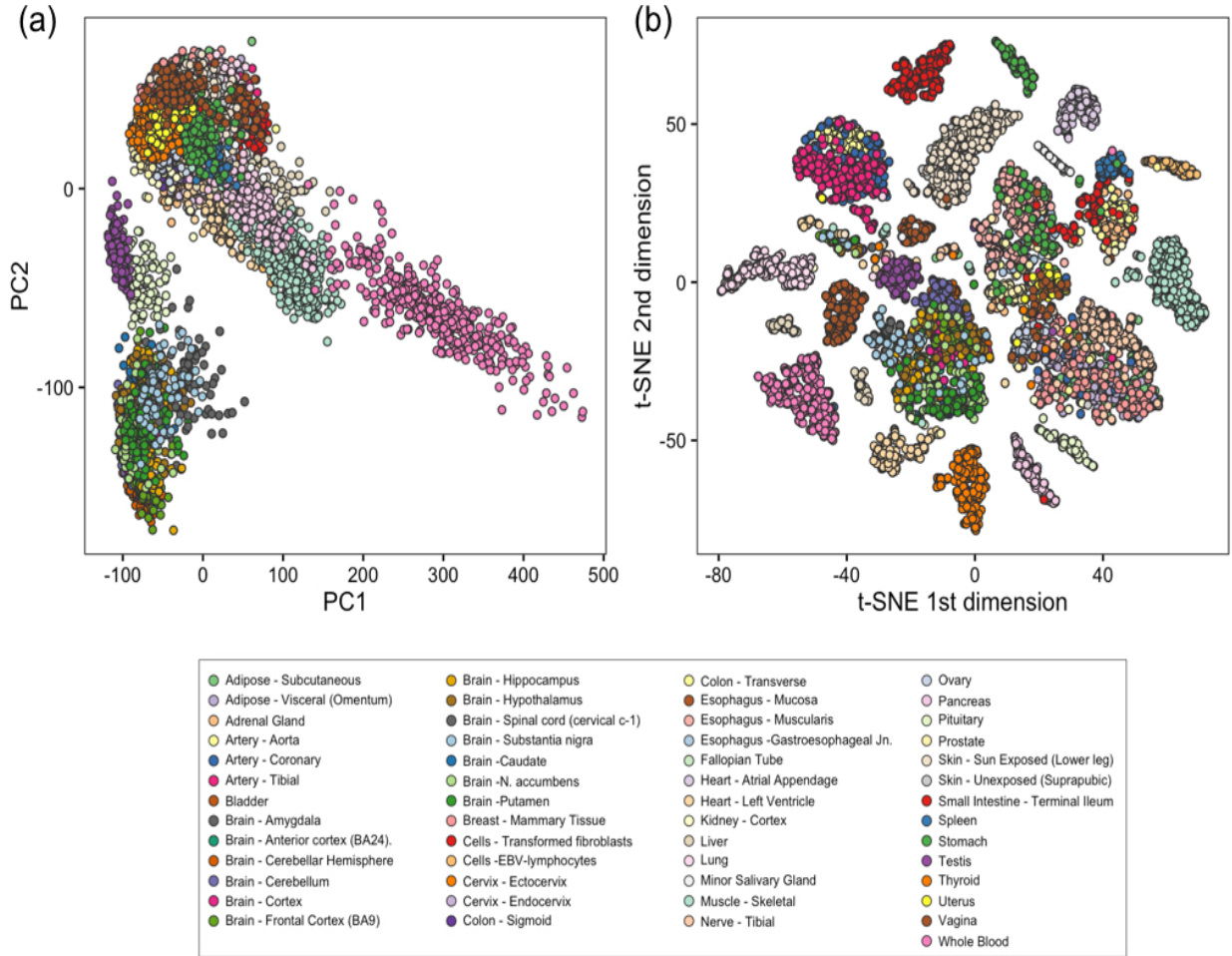


(c) hierarchy 0.0001

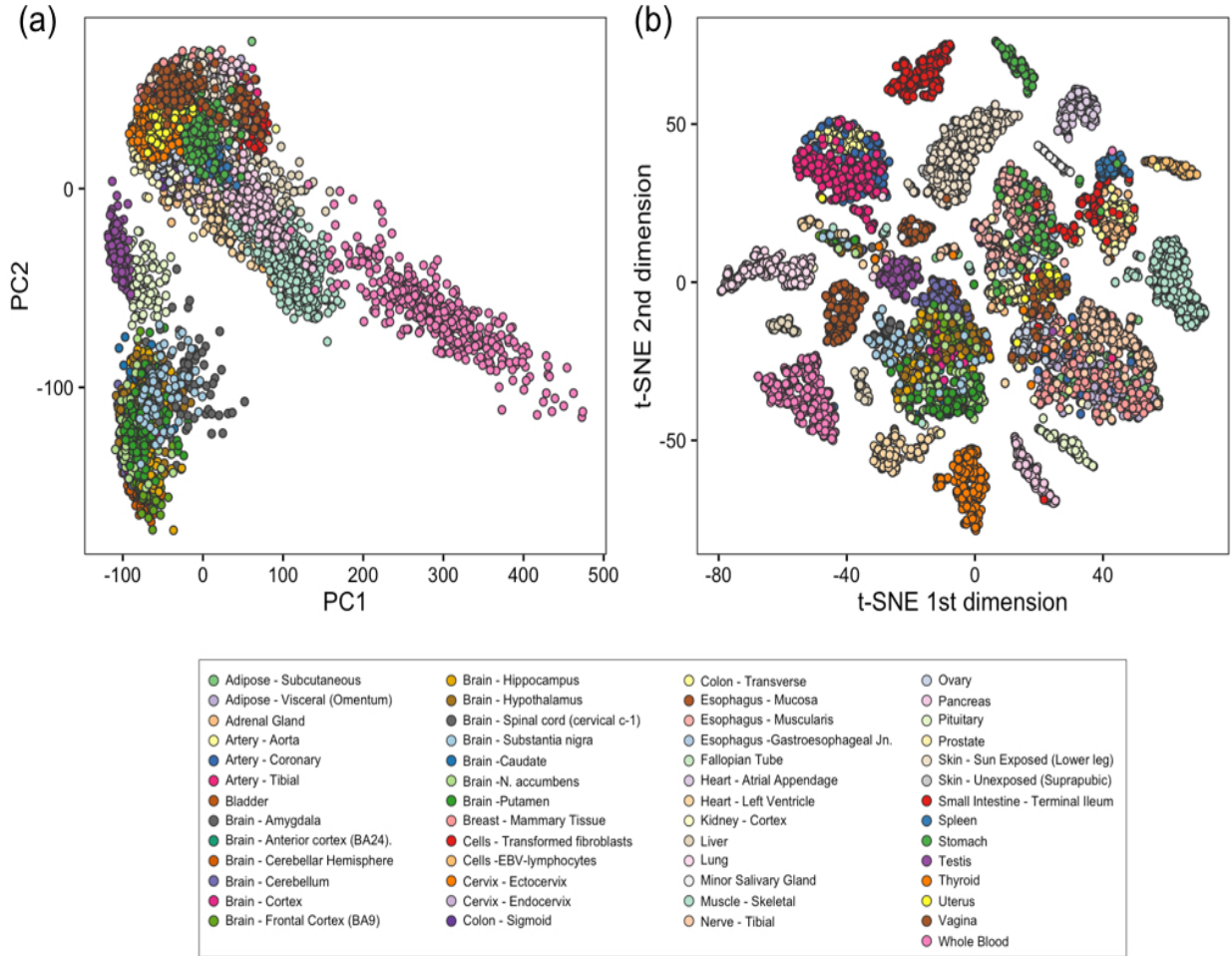


(d) GoM thin 0.0001

**S4 Fig. GTEx data visualization of all tissue samples using (a) principle component analysis and (b) t-SNE.** Samples of matching tissue types are indicated by points of matching color.



**S4 Fig. GTEx data visualization of all tissue samples using (a) principle component analysis and (b) t-SNE.** Samples of matching tissue types are indicated by points of matching color.



## 2 Supplementary tables

**S1 Table. Cluster Annotations GTEx V6 data with top driving gene summaries.**

| Cluster         | Top Driving Genes | Gene names                                    | Gene Summary  |
|-----------------|-------------------|---|---|
| 1, Royal purple | <i>NEAT1</i>      | nuclear paraspeckle assembly transcript 1     | produces a long non-coding RNA (lncRNA) transcribed from the multiple endocrine neoplasia locus, regulates genes involved in cancer progression.  |
|                 | <i>CCNL2</i>      | cyclin L2                                     | regulator of the pre-mRNA splicing process, as well as in inducing apoptosis by modulating the expression of apoptotic and antiapoptotic proteins.  |
|                 | <i>SRSF5</i>      | serine/arginine-rich splicing factor 5        | encodes proteins of serine/arginine (SR)-rich family, involved in mRNA export from the nucleus and in translation.  |
| 2, Light purple | <i>SNAP25</i>     | synaptosomal-associated protein, 25kDa        | this gene product is a presynaptic plasma membrane protein involved in the regulation of neurotransmitter release.  |
|                 | <i>FBXL16</i>     | F-box and leucine-rich repeat protein 16      | members of F-box protein family, which interact with SKP1 through the F box, and they interact with ubiquitination targets through other protein interaction domains.   |
|                 | <i>SLC17A7</i>    | neurochondrin                                 | encodes proteins expressed in neuron-rich regions; associated with the membranes of synaptic vesicles and functions in glutamate transport.   |
| 3, Red          | <i>FABP4</i>      | fatty acid binding protein 4                  | encodes the fatty acid binding protein found in adipocytes, takes part in fatty acid uptake, transport, and metabolism.   |
|                 | <i>PLIN1</i>      | perilipin 1                                   | protein encoded by this gene coats lipid storage droplets in adipocytes, thereby protecting them until they can be broken down by hormone-sensitive lipase.   |
|                 | <i>FASN</i>       | fatty acid synthase                           | catalyze the synthesis of palmitate from acetyl-CoA and malonyl-CoA, in the presence of NADPH, into long-chain saturated fatty acids.   |
| 4, Salmon       | <i>ACTG2</i>      | actin, gamma 2, smooth muscle, enteric        | involved in various types of cell motility and in the maintenance of the cytoskeleton.  |
|                 | <i>MYH11</i>      | myosin, heavy chain 11, smooth muscle         | protein encoded by this gene is a smooth muscle myosin belonging to the myosin heavy chain family, functions as a major contractile protein, converting chemical energy into mechanical energy through the hydrolysis of ATP. |
|                 | <i>SYNM</i>       | synemin                                       | protein has been found to form a linkage between desmin, which is a subunit of the IF network, and the extracellular matrix, and provides an important structural support in muscle.  |
| 5, Denim        | <i>RGS5</i>       | regulator of G-protein signaling 5            | encodes a member of the regulators of G protein signaling (RGS) family, associated with retinal arterial macroaneurysm.   |
|                 | <i>MFGE8</i>      | milk fat globule-EGF factor 8 protein         | encodes a preproprotein that is proteolytically processed to form multiple protein products, been implicated in wound healing, autoimmune disease, and cancer   |
|                 | <i>ITGA8</i>      | synemin                                       | Proteins generated mediate numerous cellular processes including cell adhesion, cytoskeletal rearrangement, and activation of cell signaling pathways.  |
| 6, Light denim  | <i>KRT10</i>      | keratin 10                                    | encodes a member of the type I (acidic) cytokeratin family, mutations associated with epidermolytic hyperkeratosis.   |
|                 | <i>KRT1</i>       | keratin 1, type II                            | specifically expressed in the spinous and granular layers of the epidermis with family member KRT10 and mutations in these genes have been associated with bullous congenital ichthyosiform erythroderma.                     |
|                 | <i>KRT2</i>       | keratin 2, type II                            | expressed largely in the upper spinous layer of epidermal keratinocytes and mutations in this gene have been associated with bullous congenital ichthyosiform erythroderma.   |
| 7, Orange       | <i>NEB</i>        | nebulin                                       | encodes nebulin, a giant protein component of the cytoskeletal matrix that coexists with the thick and thin filaments within the sarcomeres of skeletal muscle, associated with recessive nemaline myopathy.                  |
|                 | <i>MYH1</i>       | myosin, heavy chain 1, skeletal muscle, adult | a major contractile protein which converts chemical energy into mechanical energy through the hydrolysis of ATP.  |
|                 | <i>MYH2</i>       | myosin, heavy chain 2, skeletal muscle, adult | encodes a member of the class II or conventional myosin heavy chains, and functions in skeletal muscle contraction.   |



| Cluster         | Top Driving Genes | Gene names   | Gene Summary   |
|-----------------|-------------------|--|--|
| 8, Light orange | <i>FN1</i>        | fibronectin 1                                      | Fibronectin is involved in cell adhesion, embryogenesis, blood coagulation, host defense, and metastasis.  |
|                 | <i>COL1A1</i>     | collagen, type I, alpha 1                          | Mutations in this gene associated with osteogenesis imperfecta types I-IV, Ehlers-Danlos syndrome type and Classical type, Caffey Disease.   |
|                 | <i>COL1A2</i>     | collagen, type I, alpha 2                          | Mutations in this gene associated with osteogenesis imperfecta types I-IV, Ehlers-Danlos syndrome type and Classical type, Caffey Disease.   |
| 9, Green        | <i>MBP</i>        | myelin basic protein                               | major constituent of the myelin sheath of oligodendrocytes and Schwann cells in the nervous system   |
|                 | <i>GFAP</i>       | glial fibrillary acidic protein                    | encodes one of the major intermediate filament proteins of mature astrocytes, mutations causes Alexander disease.  |
|                 | <i>CARNS1</i>     | carnosine synthase 1                               | catalyzes the formation of carnosine and homocarnosine, which are found mainly in skeletal muscle and the central nervous system, respectively.  |
| 10, Light green | <i>CYP17A1</i>    | cytochrome P450 family 17 subfamily A member 1     | encodes a member of the cytochrome P450 superfamily of enzymes, mutations in this gene are associated with isolated steroid-17 alpha-hydroxylase deficiency, 20-lyase deficiency, pseudohermaphroditism, and adrenal hyperplasia.  |
|                 | <i>CYP11B1</i>    | cytochrome P450 family 11 subfamily B member 1     | The protein encoded by this gene plays a key role in the acute regulation of steroid hormone synthesis by enhancing the conversion of cholesterol into pregnenolone, associated with congenital lipoid adrenal hyperplasia.  |
|                 | <i>GKN1</i>       | gastrokin 1  | protein encoded by this gene is found to be down-regulated in human gastric cancer tissue as compared to normal gastric mucosa..   |
| 11, Turquoise   | <i>MPZ</i>        | myelin protein zero                                | specifically expressed in Schwann cells of the peripheral nervous system and encodes a type I transmembrane glycoprotein that is a major structural protein of the peripheral myelin sheath, mutations associated with autosomal dominant form of Charcot-Marie-Tooth disease type 1 and other polyneuropathies. |
|                 | <i>APOD</i>       | apolipoprotein D                                   | encodes a component of high density lipoprotein that has no marked similarity to other apolipoprotein sequences, closely associated with lipoprotein metabolism.   |
|                 | <i>PMP22</i>      | peripheral myelin protein 22                       | encodes an integral membrane protein that is a major component of myelin in the peripheral nervous system..  |
| 12, Yellow      | <i>IGHM</i>       | immunoglobulin heavy constant mu                   | IgM antibodies play an important role in primary defense mechanisms, Diseases associated with IGHM include agammaglobulinemia 1 and immunodeficiency 23.   |
|                 | <i>IGHG1</i>      | immunoglobulin heavy constant gamma 1 (G1m marker) | antigen binding functionality, diseases associated with IGHG1 include heavy chain deposition disease and chronic lymphocytic leukemia.   |
|                 | <i>IGHG2</i>      | immunoglobulin heavy constant gamma 2 (G2m marker) | antigen binding gene, diseases associated with IGHG2 include c2 deficiency.  |
| 13, Sky blue    | <i>TG</i>         | thyroglobulin                                      | thyroglobulin produced predominantly in thyroid gland, synthesizes thyroxine and triiodothyronine, associated with Graves disease and Hashimoto thyroiditis.   |
|                 | <i>PRL</i>        | prolactin 2  | encodes the anterior pituitary hormone prolactin. This secreted hormone is a growth regulator for many tissues, including cells of the immune system.  |
|                 | <i>PRM2</i>       | protamine 2  | Protamines are the major DNA-binding proteins in the nucleus of sperm.   |
| 14, Light pink  | <i>NPPA</i>       | natriuretic peptide A                              | protein encoded by this gene belongs to the natriuretic peptide family, controls extracellular fluid volume and electrolyte homeostasis, mutations Mutations associated with atrial fibrillation familial type 6.  |
|                 | <i>MYH6</i>       | myosin, heavy chain 6, cardiac muscle, alpha       | encodes the alpha heavy chain subunit of cardiac myosin, mutations cause familial hypertrophic cardiomyopathy and atrial septal defect 3   |
|                 | <i>TNNT2</i>      | protamine 2  | protein encoded by this gene is the tropomyosin-binding subunit of the troponin complex, mutations in this gene have been associated with familial hypertrophic cardiomyopathy as well as with dilated cardiomyopathy.   |

| Cluster        | Top Driving Genes | Gene names                           | Gene Summary   |
|----------------|-------------------|--------------------------------------|--|
| 15, Light gray | <i>KRT13</i>      | keratin 13, type I                   | protein encoded by this gene is a member of the keratin gene family, associated with the autosomal dominant disorder White Sponge Nevus.   |
|                | <i>KRT4</i>       | keratin 4, type II                   | protein encoded by this gene is a member of the keratin gene family, associated with White Sponge Nevus, characterized by oral, esophageal, and anal leukoplakia.  |
|                | <i>CRNN</i>       | cornulin                             | may play a role in the mucosal/epithelial immune response and epidermal differentiation.   |
| 16, Gray       | <i>SFTPB</i>      | surfactant protein B                 | an amphipathic surfactant protein essential for lung function and homeostasis after birth, mutations cause pulmonary alveolar proteinosis, fatal respiratory distress in the neonatal period.  |
|                | <i>SFTPA2</i>     | surfactant protein A2                | Mutations in this gene and a highly similar gene located nearby, which affect the highly conserved carbohydrate recognition domain, are associated with idiopathic pulmonary fibrosis.   |
|                | <i>SFTPA1</i>     | surfactant protein A1                | encodes a lung surfactant protein that is a member of C-type lectins called collectins, associated with idiopathic pulmonary fibrosis.   |
| 17, Brown      | <i>CSF3R</i>      | colony stimulating factor 3 receptor | protein encoded by this gene is the receptor for colony stimulating factor 3, a cytokine that controls the production, differentiation, and function of granulocytes, mutations a cause of Kostmann syndrome                             |
|                | <i>MMP25</i>      | matrix metalloproteinase 25          | proteins are involved in the breakdown of extracellular matrix in normal physiological processes, such as embryonic development, reproduction, and tissue remodeling, as well as in disease processes, such as arthritis and metastasis. |
|                | <i>IL1R2</i>      | interleukin 1 receptor type 2        | protein encoded by this gene is a cytokine receptor that belongs to the interleukin 1 receptor family.   |
| 18, Purple     | <i>PRSS1</i>      | protease, serine 1                   | secreted by pancreas, associated with pancreatitis   |
|                | <i>CPA1</i>       | carboxypeptidase A1                  | secreted by pancreas, linked to pancreatitis and pancreatic cancer   |
|                | <i>PNLIP</i>      | pancreatic lipase                    | encodes a carboxyl esterase that hydrolyzes insoluble, emulsified triglycerides, and is essential for the efficient digestion of dietary fats. This gene is expressed specifically in the pancreas.                                      |
| 19, Pink       | <i>HBB</i>        | hemoglobin, beta                     | mutant beta globin causes sickle cell anemia, absence of beta chain/ reduction in beta globin leads to thalassemia.  |
|                | <i>HBA2</i>       | hemoglobin, alpha 2                  | deletion of alpha genes may lead to alpha thalassemia.   |
|                | <i>HBA1</i>       | hemoglobin, alpha 1                  | deletion of alpha genes may lead to alpha thalassemia.   |
| 20, Dark gray  | <i>ALB</i>        | albumin                              | functions primarily as a carrier protein for steroids, fatty acids, and thyroid hormones and plays a role in stabilizing extracellular fluid volume.   |
|                | <i>HP</i>         | haptoglobin                          | encodes a preproprotein, which subsequently produces haptoglobin, linked to diabetic nephropathy, Crohn's disease, inflammatory disease behavior and reduced incidence of Plasmodium falciparum malaria.                                 |
|                | <i>FGB</i>        | fibrinogen beta chain                | protein encoded by this gene is the beta component of fibrinogen, mutations may lead to several disorders, including afibrinogenemia, dysfibrinogenemia, hypodysfibrinogenemia etc.  |

**S2 Table. Cluster Annotations GTEx V6 Brain data with top driving gene summaries.**

| Cluster          | Top Driving Genes | Gene names   | Gene Summary  |
|------------------|-------------------|--|---|
| 1, Royal blue    | <i>CLU</i>        | clusterin  | protein encoded by this gene is a secreted chaperone that can under some stress conditions also be found in the cell cytosol, also involved in cell death, tumor progression, and neurodegenerative disorders.                              |
|                  | <i>OXT</i>        | oxytocin/neurophysin I pre-propeptide                                      | encodes a precursor protein that is processed to produce oxytocin and neurophysin I, involved in contraction of smooth muscle during parturition and lactation, cognition, tolerance, adaptation and complex sexual and maternal behaviour. |
|                  | <i>GLUL</i>       | glutamate-ammonia ligase   | catalyzes the synthesis of glutamine from glutamate and ammonia in an ATP-dependent reaction, associated with congenital glutamine deficiency, and overexpression of this gene was observed in some primary liver cancer samples.           |
| 2, Turquoise     | <i>ENC1</i>       | ectodermal-neural cortex 1   | plays a role in the oxidative stress response as a regulator of the transcription factor Nrf2, may play role in malignant transformation.   |
|                  | <i>NCALD</i>      | neurocalcin delta  | encodes a member of the neuronal calcium sensor (NCS), a regulator of G protein-coupled receptor signal transduction.   |
|                  | <i>YWHAH</i>      | tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein eta | mediate signal transduction by binding to phosphoserine-containing proteins, associated with early-onset schizophrenia and psychotic bipolar disorder.  |
| 3, Lime green    | <i>PKD1</i>       | polycystin 1, transient receptor potential channel interacting             | functions as a regulator of calcium permeable cation channels and intracellular calcium homeostasis. It is also involved in cell-cell/matrix interactions and may modulate G-protein-coupled signal-transduction pathways.                  |
|                  | <i>CBLN3</i>      | cerebellin 3 precursor   | contain a cerebellin motif and C-terminal C1q signature domain that mediates trimeric assembly of atypical collagen complexes   |
|                  | <i>CHGB</i>       | chromogranin B   | encodes a tyrosine-sulfated secretory protein abundant in peptidergic endocrine cells and neurons. This protein may serve as a precursor for regulatory peptides.   |
| 4, Red           | <i>PPP1R1B</i>    | protein phosphatase 1 regulatory inhibitor sub- unit 1B                    | encodes a bifunctional signal transduction molecule, may serve as a therapeutic target for neurologic and psychiatric disorders.  |
|                  | <i>RGS14</i>      | regulator of G-protein signaling 14  | attenuates the signaling activity of G-proteins, increases the rate of conversion of the GTP to GDP.  |
|                  | <i>NCDN</i>       | neurochondrin  | encodes a leucine-rich cytoplasmic protein, essential for spatial learning processes.   |
| 5, Yellow orange | <i>MBP</i>        | myelin basic protein   | protein encoded is a major constituent of the myelin sheath of oligodendrocytes and Schwann cells in the nervous system.  |
|                  | <i>GFAP</i>       | glial fibrillary acidic protein  | encodes major intermediate filament proteins of mature astrocytes, a marker to distinguish astrocytes during development, mutations in this gene cause Alexander disease, a rare disorder of astrocytes in central nervous system.          |
|                  | <i>TF</i>         | transferrin  | transport iron from the intestine, reticuloendothelial system, and liver parenchymal cells to all proliferating cells in the body, involved in the removal of certain organic matter and allergens from serum.                              |
| 6, Yellow        | <i>IQGAP1</i>     | IQ motif containing GTPase activating protein 1                            | interacts with components of the cytoskeleton, with cell adhesion molecules, and with several signaling molecules to regulate cell morphology and motility.   |
|                  | <i>A2M</i>        | alpha-2-macroglobulin  | inhibits many proteases, including trypsin, thrombin and collagenase. A2M is implicated in Alzheimer disease (AD) due to its ability to mediate the clearance and degradation of A-beta, the major component of beta-amyloid deposits.      |
|                  | <i>C3</i>         | complement component 3   | plays a central role in the activation of complement system, associated with atypical hemolytic uremic syndrome and age-related macular degeneration in human patients.   |

**S3 Table. Deng et al (2014) Cluster GO annotations of top driving genes.**

|    | go.id      | name  | significant   |
|----|------------|---|---|
| 1  | GO:0007276 | gamete generation                             | BCL2L10; GDF9; NOBOX; PABPC1L; RGS2; CREB3L4; RNF114; BMP15; PTTG1; TDRD12; WEE2; SPIN1; DAZL               |
| 2  | GO:0007292 | female gamete generation                      | GDF9; BCL2L10; PABPC1L; BMP15; WEE2; DAZL; NOBOX  |
| 3  | GO:0048609 | multicellular organismal reproductive process | GDF9; NOBOX; PABPC1L; BCL2L10; BMP15; CREB3L4; TGFB2; RNF114; RGS2; PTTG1; TDRD12; WEE2; SPIN1; DAZL        |
| 4  | GO:0032504 | multicellular organism reproduction           | GDF9; NOBOX; PABPC1L; BCL2L10; BMP15; CREB3L4; TGFB2; RNF114; RGS2; PTTG1; TDRD12; WEE2; SPIN1; DAZL        |
| 5  | GO:0019953 | sexual reproduction                           | BCL2L10; GDF9; NOBOX; PABPC1L; RGS2; CREB3L4; RNF114; BMP15; PTTG1; TDRD12; WEE2; SPIN1; DAZL               |
| 6  | GO:0044702 | single organism reproductive process          | GDF9; NOBOX; PABPC1L; BCL2L10; BMP15; CREB3L4; TGFB2; CASP8; RNF114; RGS2; PTTG1; TDRD12; WEE2; SPIN1; DAZL |
| 7  | GO:0048477 | oogenesis                                     | WEE2; GDF9; NOBOX; PABPC1L; DAZL  |
| 8  | GO:0044703 | multi-organism reproductive process           | BCL2L10; GDF9; NOBOX; PABPC1L; RGS2; CREB3L4; RNF114; BMP15; PTTG1; TDRD12; WEE2; SPIN1; DAZL               |
| 9  | GO:0048599 | oocyte development                            | WEE2; GDF9; PABPC1L; DAZL   |
| 10 | GO:0009994 | oocyte differentiation                        | WEE2; GDF9; PABPC1L; DAZL   |
| 11 | GO:0051321 | meiotic cell cycle                            | H1FOO; WEE2; TDRD12; SPIN1; PTTG1; DAZL   |
| 12 | GO:0001556 | oocyte maturation                             | WEE2; PABPC1L; DAZL   |
| 13 | GO:0006306 | DNA methylation                               | TDRD12; H1FOO; TET3; ZFP57  |
| 14 | GO:0051302 | regulation of cell division                   | TGFB2; PTTG1; TXNIP; WEE2; CHEK1; DAZL  |
| 15 | GO:0060255 | regulation of macromolecule metabolic process | TGFB2; NOBOX; BPGM; UBE2D3; NFYA; CASP8; BMP15; TXNIP; TDRD12; GDF9; BCL2L10                                |

**S3 Table continued. Deng et al (2014) Cluster 2 (magenta) top GO annotations.**

|   | go.id      | name                               | significant  |
|---|------------|------------------------------------|--|
| 1 | GO:0016604 | nuclear body                       | YTHDC1; RBM8A; CDK12; PSME4; PPP1R8; HIPK1; TOPORS |
| 2 | GO:0005814 | centriole                          | SFI1; PLK2; ROCK1; TOPORS                          |
| 3 | GO:0044450 | microtubule organizing center part | SFI1; PLK2; ROCK1; TOPORS                          |

**S3 Table continued. Deng et al (2014) Cluster 3 (yellow) top GO annotations.**

|    | go.id      | name                                     | significant  |
|----|------------|--|--|
| 1  | GO:0044428 | nuclear part                             | MAD2L2; SMARCC1; PPRC1; SLU7; NFYB; TOR1B; MIOS; NR1H3; POLR3K   |
| 2  | GO:0031981 | nuclear lumen                            | MAD2L2; SMARCC1; PPRC1; SLU7; NFYB; POLR1E; MIOS; POLR3K; XPO1   |
| 3  | GO:0070013 | intracellular organelle lumen            | MAD2L2; SMARCC1; PPRC1; SLU7; NFYB; POLR1E; MIOS; POLR3K; XPO1; DNTTIP2; ZBTB10; ZBTB17                      |
| 4  | GO:0043233 | organelle lumen                          | MAD2L2; SMARCC1; PPRC1; SLU7; NFYB; POLR1E; MIOS; POLR3K; XPO1   |
| 5  | GO:0005730 | nucleolus                                | XPO1; DNTTIP2; ESF1; WDR43; ZDHHC7; HEATR1; POLR1E; DDX24; POLR3K  |
| 6  | GO:0005634 | nucleus                                  | MAD2L2; SMARCC1; PPRC1; SLU7; NFYB; TOR1B; MIOS; NR1H3; EIF5B; POLR3K  |
| 7  | GO:0044446 | intracellular organelle part             | MAD2L2; PTDSS2; SMARCC1; KLHL21; TOR1B; PPRC1; SLU7; NFYB; SLC25A36; ECE2                                    |
| 8  | GO:0005654 | nucleoplasm                              | MAD2L2; SMARCC1; PPRC1; SLU7; NFYB; POLR1E; MIOS; POLR3K; XPO1; ZBTB10; ZBTB17                               |
| 9  | GO:0003723 | RNA binding                              | PPRC1; EIF5B; XPO1; DNTTIP2; WDR43; DDX10; EIF3C; BCLAF1; EBNA1BP2; RARS                                     |
| 10 | GO:0003676 | nucleic acid binding                     | SMARCC1; PPRC1; SLU7; NFYB; POLR1E; EIF5B; POLR3K; XPO1; DNTTIP2   |
| 11 | GO:0043231 | intracellular membrane-bounded organelle | MAD2L2; PTDSS2; SMARCC1; TOR1B; PPRC1; SLU7; NFYB; ESF1; ECE2; LMAN1L  |
| 12 | GO:0043229 | intracellular organelle                  | MAD2L2; PTDSS2; SMARCC1; KLHL21; TOR1B; PPRC1; ARRDC1; SLU7; NFYB; ESF1; ECE2                                |
| 13 | GO:0005874 | microtubule                              | WDR43; KLHL21; HAUS6; CENPE; TEK2; RACGAP1; WDR81; BCL2L11; KIF20B   |
| 14 | GO:0044822 | poly(A) RNA binding                      | WDR43; DNTTIP2; ESF1; NXF1; DDX10; HEATR1; EIF3C   |
| 15 | GO:0044424 | intracellular part                       | MAD2L2; PTDSS2; SMARCC1; KLHL21; TOR1B; PPRC1; SNAPC4; POLR3K; ARRDC1; SLU7; NFYB; ESF1; WDR43; ECE2; LMAN1L |

**S3 Table continued. Deng et al (2014) Cluster 4 (green) top GO annotations.**

|    | go.id      | name                                      | significant   |
|----|------------|---|---|
| 1  | GO:0005829 | cytosol                                   | PARG; UAP1; PSMB10; TCEB1; RPLP0; EIF5; CNBP; RPS3; PSAT1; AACS; PMM1; EXOSC7; EIF3I; SET; BHMT; BHMT2                      |
| 2  | GO:0044444 | cytoplasmic part                          | PARG; UAP1; PSMB10; TCEB1; HSPA8; SERINC1; EIF5; CNBP; RPS3; PSAT1; GPD2; AACS; GPR137B; STIP1; PMM1; EXOSC7; VPRED3; PEX16 |
| 3  | GO:0055131 | C3HC4-type RING finger domain binding     | HSPA8; PINK1; DNAJA1  |
| 4  | GO:1901575 | organic substance catabolic process       | PSMB10; TCEB1; RPLP0; RPS3; GPD2; PINK1; EXOSC7; ALLC; BHMT; HSP90AB1; RPL13A; ATG7; CUL5; UBXN1; ZMPSTE24                  |
| 5  | GO:0000151 | ubiquitin ligase complex                  | DNAJA1; RNF7; UBE2C; HSPA8; FBXO15; SUGT1; DCAF4; CUL5; FBXL20  |
| 6  | GO:0072655 | protein localization to mitochondrion     | TIMM17A; BNIP3L; ARIH2; PEMT; SFN; PINK1; HSP90AA1; TIMM23  |
| 7  | GO:1901564 | organonitrogen compound metabolic process | PSMB10; RPLP0; SERINC1; EIF5; BHMT2; PINK1; EIF3I; ALLC; BHMT; MRPL22; RPL13A; ATG7; NUDT9; VNN1; CTSA; HK1                 |
| 8  | GO:0005737 | cytoplasm                                 | PARG; UAP1; PSMB10; TCEB1; HSPA8; SERINC1; EIF5; CNBP; RPS3; PSAT1; GPD2; AACS; GPR137B; STIP1; PMM1; EXOSC7                |
| 9  | GO:0044265 | cellular macromolecule catabolic process  | EXOSC7; SUMO2; BNIP3L; ARIH2; PSMB10; TCEB1; RPLP0; UBXN1; HSP90AB1; RPL13A; RPS3; RNF7; PINK1                              |
| 10 | GO:0023026 | MHC class II protein complex binding      | HSP90AB1; HSP90AA1; HSPA8   |
| 11 | GO:0051082 | unfolded protein binding                  | DNAJA1; PTGES3; HSPA8; HSP90AB1; HSP90AA1; NPM1   |
| 12 | GO:0009056 | catabolic process                         | PSMB10; TCEB1; RPLP0; RPS3; GPD2; PINK1; EXOSC7; ALLC; WDR45; HSP90AB1; RPL13A  |
| 13 | GO:0009057 | macromolecule catabolic process           | EXOSC7; SUMO2; BNIP3L; ARIH2; PSMB10; TCEB1; RPLP0; AZIN1; UBXN1; HSP90AB1; RPL13A  |
| 14 | GO:0044248 | cellular catabolic process                | PSMB10; TCEB1; SUMO2; RPS3; GPD2; PINK1; EXOSC7; ALLC; WDR45; HSP90AB1  |
| 15 | GO:0006626 | protein targeting to mitochondrion        | TIMM17A; BNIP3L; ARIH2; PEMT; PINK1; HSP90AA1; TIMM23   |

**S3 Table continued. Deng et al (2014) Cluster 5 (purple) top GO annotations.**

|    | go.id      | name                                     | significant   |
|----|------------|--|---|
| 1  | GO:0044710 | single-organism metabolic process        | PCK2; SAT1; EPHX2; NFATC4; CKB; PRDX6; MSH2; EPHA4; PROS1; PDGFRA; PRDX1; UBE2L6; POGLUT1; FABP5; AKAP12; TDGF1; FBP2; SOX2 |
| 2  | GO:0006950 | response to stress                       | EPHX2; NFATC4; PRDX6; MSH2; EPHA4; PROS1; PDGFRA; PRDX1; UBE2L6; FABP5; TDGF1; SOX2   |
| 3  | GO:0065010 | extracellular membrane-bounded organelle | PCK2; EPHX2; MFGE8; CKB; PRDX6; PROS1; PRDX1; POGLUT1; FABP5; FBP2; TRAP1; PLOD2; DHRS4                                     |
| 4  | GO:0070062 | extracellular exosome                    | PCK2; EPHX2; MFGE8; CKB; PRDX6; PROS1; PRDX1; POGLUT1; FABP5; FBP2; TRAP1; PLOD2; DHRS4; MARCKS; DPP4; PRKCI; RAC2; IDH1    |
| 5  | GO:0043230 | extracellular organelle                  | PCK2; EPHX2; MFGE8; CKB; PRDX6; PROS1; PRDX1; POGLUT1; FABP5; FBP2; TRAP1; PLOD2; DHRS4; MARCKS; DPP4                       |
| 6  | GO:1903561 | extracellular vesicle                    | PCK2; EPHX2; MFGE8; CKB; PRDX6; PROS1; PRDX1; POGLUT1; FABP5; FBP2; TRAP1; PLOD2; DHRS4; MARCKS; DPP4; PRKCI                |
| 7  | GO:0042221 | response to chemical                     | EPHX2; NFATC4; MFGE8; PRDX6; EPHA4; PROS1; PDGFRA; PRDX1; UBE2L6; TDGF1; SOX2   |
| 8  | GO:0031988 | membrane-bounded vesicle                 | PCK2; EPHX2; MFGE8; CKB; PRDX6; PROS1; PRDX1; POGLUT1; FABP5; FBP2; TRAP1; PLOD2; DHRS4; SPARC                              |
| 9  | GO:0031982 | vesicle                                  | PCK2; EPHX2; MFGE8; CKB; PRDX6; PROS1; PRDX1; POGLUT1; FABP5; FBP2; TRAP1; PLOD2; DHRS4; SPARC                              |
| 10 | GO:0001525 | angiogenesis                             | SAT1; PDGFRA; BMP4; NFATC4; MFGE8; FN1; MEIS1; SPARC; COL4A2; COL4A1; FGF10; TDGF1  |
| 11 | GO:0048514 | blood vessel morphogenesis               | SAT1; PDGFRA; BMP4; NFATC4; MFGE8; FN1; ZFP36L1; MEIS1; SPARC; COL4A2; COL4A1; FGF10; TDGF1                                 |
| 12 | GO:0001944 | vasculature development                  | SAT1; PDGFRA; BMP4; NFATC4; MFGE8; FN1; ZFP36L1; MEIS1; PDPN; SPARC; COL4A2; COL4A1; FGF10; TDGF1                           |
| 13 | GO:0006979 | response to oxidative stress             | TAT; PDGFRA; BMP4; ETV5; TRAP1; PRDX6; IDH1; PARP1; AQP8; PRDX1; CRYGD  |
| 14 | GO:0009725 | response to hormone                      | PRKCI; GJA1; PDGFRA; BMP4; MFGE8; TAT; PLOD2; SPP1; IDH1  |
| 15 | GO:0030198 | extracellular matrix organization        | PDGFRA; BMP4; JAM2; FN1; PLOD2; SPARC; SPP1; COL4A2; COL4A1; SERPINH1; DPP4   |

**S3 Table continued. Deng et al (2014) Cluster 6 (orange) top GO annotations.**

|    | go.id      | name                                     | genes  |
|----|------------|--|--|
| 1  | GO:0065010 | extracellular membrane-bounded organelle | MYH10; SLC2A3; GM2A; TSPAN8; ACTG1; SDC4; TINAGL1; CRYAB; MSN; FABP3; PDZK1IP1; PRSS8; S100A11; DAB2; KRT8; LCP1; UGP2 |
| 2  | GO:0070062 | extracellular exosome                    | MYH10; SLC2A3; GM2A; TSPAN8; ACTG1; SDC4; TINAGL1; CRYAB; MSN; FABP3; PDZK1IP1; PRSS8; S100A11; DAB2; KRT8; LCP1; UGP2 |
| 3  | GO:0043230 | extracellular organelle                  | MYH10; SLC2A3; GM2A; TSPAN8; ACTG1; SDC4; TINAGL1; CRYAB; MSN; FABP3; PDZK1IP1; PRSS8; S100A11                         |
| 4  | GO:1903561 | extracellular vesicle                    | MYH10; SLC2A3; GM2A; TSPAN8; ACTG1; SDC4; TINAGL1; CRYAB; MSN; FABP3; PDZK1IP1; PRSS8; S100A11; DAB2; KRT8             |
| 5  | GO:0031988 | membrane-bounded vesicle                 | MYH10; SLC2A3; GM2A; TSPAN8; ACTG1; TMSB4X; SDC4; TINAGL1; CRYAB; MSN; FABP3; PDZK1IP1; PRSS8; S100A11; DAB2           |
| 6  | GO:0031982 | vesicle                                  | MYH10; SLC2A3; GM2A; TSPAN8; ACTG1; TMSB4X; SDC4; TINAGL1; CRYAB; MSN; FABP3; PDZK1IP1; PRSS8; S100A11; DAB2; KRT8     |
| 7  | GO:0008092 | cytoskeletal protein binding             | MYH10; TPM4; TMSB4X; CRYAB; MSN; TMSB10; FABP3; NDRG1; CALM1; FMNL2; MYH9; CAP1; TPM1; CDH1                            |
| 8  | GO:0015629 | actin cytoskeleton                       | MYH10; CLIC4; MYH9; MYL12B; WDR1; CNN2; ARPC2; AHNAK; ACTN4; CRYAB; CAP1; TPM1; DSTN; ARPC5; TPM4                      |
| 9  | GO:0003779 | actin binding                            | MYH10; TPM4; WDR1; CNN2; FMNL2; ARPC2; MYH9; CAP1; TPM1  |
| 10 | GO:0048468 | cell development                         | MYH10; CAPG; ACTG1; WDR1; CNN2; FMNL2; MYH9; ACTN4; SDC4; CAP1; TPM1; DSTN   |
| 11 | GO:0030036 | actin cytoskeleton organization          | MYH10; CAPG; ACTG1; WDR1; CNN2; FMNL2; MYH9; ACTN4; SDC4; CAP1; TPM1   |
| 12 | GO:0032432 | actin filament bundle                    | MYH10; TPM4; MYL12B; CNN2; MYH9; CRYAB; TPM1; ACTN4; LCP1  |
| 13 | GO:0005912 | adherens junction                        | TJP2; MYH9; ACTG1; CNN2; ARPC2; AHNAK; ACTN4; SDC4   |
| 14 | GO:0070161 | anchoring junction                       | TJP2; MYH9; ACTG1; CNN2; ARPC2; AHNAK; ACTN4; SDC4   |
| 15 | GO:0005925 | focal adhesion                           | MYH9; ACTG1; CNN2; ARPC2; AHNAK; ACTN4; SDC4; CAP1; ARPC5  |