PhD degree in Systems Medicine (curriculum in Human Genetics/Molecular Oncology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples “Federico II”

Settore disciplinare: BIO/13

SYSMET: SYStems biology of Membrane Trafficking

MARIO FAILLI

TIGEM, Pozzuoli

Matricola n. R10724

Supervisor:

Prof. Maria Antonietta De Matteis

TIGEM, Pozzuoli

Internal Co-supervisor:

Prof. Diego Di Bernardo

TIGEM, Pozzuoli

Coordinator:

Prof. Andrea Ballabio

TIGEM, Pozzuoli

External Co-supervisor:

Prof. Roderic Guigó

CRG, Barcelona

Anno accademico 2016-2017

To Antonella and Francesco,

who made all this possible.

List of contents

ABSTRACT

Studies on membrane trafficking have expanded massively over the last 40 years. During this time, research has led to an understanding of the molecular mechanisms underlying membrane trafficking pathways, providing crucial insights into several fundamental events. Although we have gained detailed knowledge about the molecular organization of membrane trafficking machineries there is a lack of a global view of its function, organization and regulation. In addition, many genes of the membrane trafficking machinery have been associated with diseases. In the majority of cases, disease manifestation is tissue-specific despite the ubiquitous expression of the causal gene. Explanations for this phenomenon may be found either in the specific requirements and demands of a cell within a given tissue or in differences in the expression of disease gene interactors.

The main aim of this project was to delineate sets of co-expressed membrane trafficking genes and proteins (membrane trafficking modules; MTMs) across tissues. For this purpose we curated a list of 1,261 genes that have been described as part of membrane trafficking machineries in different cellular organelles, around which we have developed a bioinformatics pipeline in order to address two specific questions:

a) are membrane trafficking genes organized in MTMs, defined as communities of co-expressed genes, and are they associated with general cellular functions?

b) do disease genes have specific membrane-trafficking co-expressed communities in those tissues that are affected by the disease?

To address these questions we used data from the Genotype-Tissue Expression (GTEx) project, a catalog of human tissue-specific gene expression patterns

obtained from “non-diseased” tissues sampled from recently deceased human donors.

With regards to the first question, we analyzed the expression patterns of the trafficking genes in twenty-five different tissues and used weighted correlation network analysis (WGCNA) to derive highly preserved MTMs. We have analyzed in more detail one that includes genes apparently involved in collagen secretion.

Instead for the second question we applied differential co-expression before the WGCNA to generate tissue-specific MTMs to understand how specific membrane trafficking gene modules might be organized in human tissues.

Chapter 1

Introduction:

Eucaryotic cells have half of their internal volume occupied by a complex system of intracellular membranes that forms functionally specialized compartments: the cellular organelles. Each organelle contains a distinct set of proteins, which mediates its unique function, and since the synthesis of the majority of them begins on ribosomes in the cytosol, non-cytosolic proteins usually have sorting signals, at the N-terminus of the amino acid sequence, that direct their delivery to locations outside the cytosol (i.e. nucleus, endoplasmic reticulum (ER), mitochondria or peroxisomes).

Sorting signals can also direct the transport of proteins from the ER to other membrane-enclosed compartments or between these compartments, tracing the principal ways of membrane transport: the exocytic or secretory pathway, that delivers newly synthesized proteins to either the plasma membrane (PM) or the extracellular space, and the endocytic pathway, that delivers PM components to internal compartments called endosomes, from where they can be recycled to the same or different regions of the PM or delivered to lysosomes for degradation.

Hence, membrane trafficking can be defined as the process by which proteins and other macromolecules are transported between chemically distinct organelles that collectively comprise the exocytic and endocytic pathways.

1.1 Membrane trafficking players

1.1.1 Proteins involved in protein folding

Analysis of the eukaryotic proteome indicate that a third of translated proteins enter the secretory pathway (Dancourt and Barlowe, 2010). The pathway starts with the translocation of the nascent polypeptides across or into the ER membrane

where they must fold before being transported to the Golgi apparatus or another destination; molecular chaperones and folding enzymes assist the correct folding of these proteins in the ER.

The ER contains two major chaperone systems dedicated to the proper translocation, maturation and sorting of the nascent chains in the early secretory pathway: the classical chaperone and the carbohydrate-binding chaperone systems. The classical chaperone system is found in almost all cellular locations and generally involves heat shock proteins (Hsp) that bind directly to the polypeptide chain. In contrast, carbohydrate-binding chaperones, such as calnexin, calreticulin and lectin chaperones, are ER-specific and interact with the hydrophilic glycan modification of the protein (see section 1.2.2). These two systems co-work to ensure that protein flux through the ER is adequately maintained for the large variety of proteins that traverse the secretory pathway (Braakman and Hebert, 2013).

In addition to chaperones, two subclasses of enzymes catalyze rate-limiting reactions during folding: the disulfide oxidoreductases, known as protein disulfide isomerases (PDIs), and the protein proline isomerases (PPIs).

For proteins that contain multiple cysteine, PDIs catalyze both oxidation (S—S) and reduction (-SH) of disulfide bonds allowing the protein to attain the pattern of disulfide bonds that is compatible with its stably folded conformation. It also happens that a protein needs a *cis*-proline to reach its folded state. In this case, the PPIs catalyze the isomerization of proline residues from *trans* to *cis* conformation, as these residues are inserted by the ribosome only in the *trans* conformation.

1.1.2 Vesicular transport effectors

Protein transfer between organelles is mainly mediated by transport vesicles. Each vesicle transport reaction can be divided into four essential steps, budding,

transport, tethering, and fusion, tightly maintained and regulated by specific trafficking components (Cai et al., 2007) (Fig. 1.1).

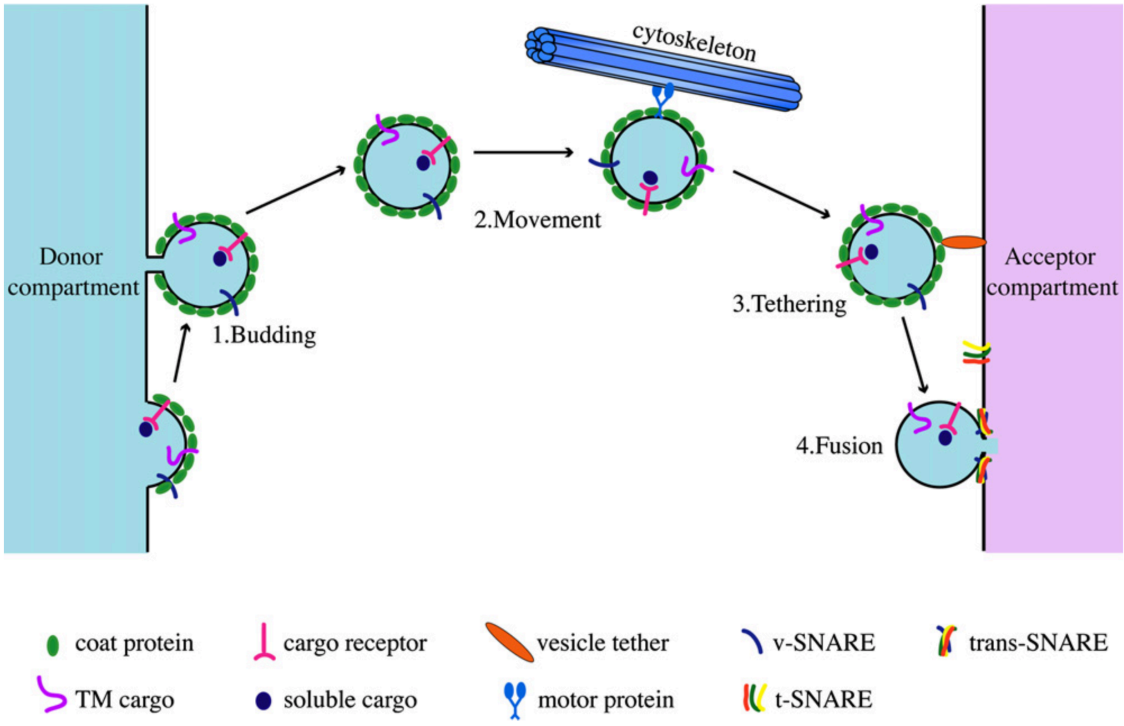


Figure 1.1 Steps in Vesicle Transport. (1) Budding: coat proteins are recruited onto the donor membrane to induce the formation of a vesicle. Cargo receptors and SNAREs are incorporated into the budding vesicle by binding to coat subunits. (2) Movement: the vesicle moves toward the acceptor compartment by diffusion or with the aid of a cytoskeletal track. (3) Tethering: tethering factors mediate the initial attachment of transport vesicle to the acceptor membrane. (4) Fusion: the vesicle-associated SNARE and the SNARE on the acceptor membrane assemble into a four-helix bundle (trans-SNARE complex), which drives membrane fusion and the delivery of cargo (Adapted from Cai et al., 2007).

1.1.2.1 Coat and cargo receptor proteins

Coat proteins constitute the trafficking machinery involved in the budding of transport vesicles. Once recruited from the cytosol to membranes, they assemble in basketlike structures that mold flat membrane patches into round buds, shaping the nascent vesicles.

The other important function of these proteins is the selective recognition of sorting signals present in the cytosolic domain of trans-membrane (TM) cargo proteins, including cargo receptors that span the membrane to efficiently recruit soluble cargo proteins to nascent vesicles (Gomez-Navarro and Miller, 2016).

There are three well-characterized types of coated vesicles, distinguished by their coat proteins that mediate vesicle budding and cargo selection at different stages of the exocytic and endocytic pathways: (I) clathrin-coated; (II) COPI-coated and (III) COPII-coated vesicles.

(I) Clathrin-coated vesicles (CCVs) bud from post-Golgi locations including the PM, the trans-Golgi network (TGN), and endosomes. The major protein component is clathrin whose subunit consists of three clathrin heavy chains and three light chains that together form a three-legged structure called triskelion. Clathrin triskelion cannot bind to membrane or cargo directly but it uses adaptor proteins to anchor to membrane and bind to other membrane-attached triskelia, forming a basketlike framework of hexagons and pentagons that pulls the membrane into a bud. Adaptor proteins constitute the inner layer of the coat in contact with various TM proteins, including cargo receptors; there exist several types of them, each specific for a different set of cargo receptors.

(II) COPI vesicles bud from the ER-Golgi intermediate compartment (ERGIC), and from the Golgi apparatus. The COPI coat is composed of heptameric complexes called coatomers composed of two subcomplexes: a trimeric complex composed of COPA/COPB2/COPE (RET1/SEC27/SEC28 in yeast) that forms the outer layer of the COPI coat, and a tetrameric complex composed of COPG (present in two isoforms COPG1 and COPG2)/ARCN1/COPZ (also present in two isoforms COPZ1

and COPZ2)/COPB1 (SEC21/RET2/RET3/SEC26 in yeast) that forms the inner core of the coat (Szul and Sztul, 2011).

(III) COPII vesicles bud from sub-domains of the ER, named ER exit sites (ERES) and are composed of five rapidly cycling proteins: the small GTPase Sar1, the inner coat (Sec23/Sec24) and the outer layer (Sec13/31). Sar1-GTP (guanosine triphosphate) together with Sec23/Sec24 constitute the so-called "pre-budding complex", which has the appearance of a bow tie with one side corresponding to Sec23 and the other to Sec24 (Bi et al., 2002). Sec23 makes direct contact with Sar1-GTP while Sec24 participates in cargo recognition. The Sec13/31 assembly unit is a heterotetramer comprising Sec13/Sec31-Sec31/Sec13, the "architectural core" of which is organized to form a long rod. Twenty-four copies of this rod assembly are required to form the COPII cuboctahedron (Stagg et al., 2006). Additional regulatory proteins participate in COPII assembly, including Sec16, a putative scaffold protein (Espenshade et al., 1995), and Sec12, a guanine nucleotide exchange factor (GEF) for Sar1 (Barlowe and Schekman, 1993).

A schematic representation of the exocytic and endocytic pathways, and of the inner and outer layers of both the COPI and COPII complexes, are provided in Figure 1.2A and B.

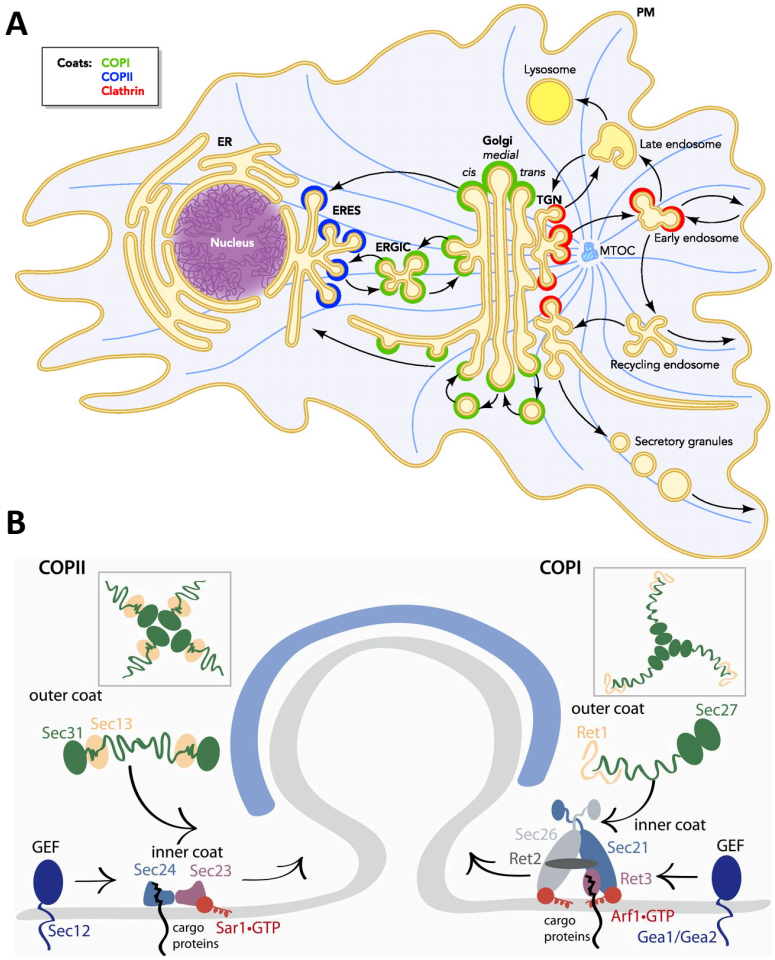


Figure 1.2 Intracellular transport pathways and coat assembly formation. (A) The scheme depicts the compartments of the secretory and endosomal pathways. Arrows indicate transport steps. Colors indicate the known coats: COPII (blue), COPI (green), and clathrin (red). Secretory cargos are synthesized in the ER, exit the ER at ERES in COPII-coated vesicles, and are transported to ERGIC. Cargos are sorted from ERGIC into anterograde carriers that move them to Golgi. After passage through the Golgi, cargos are sorted at the TGN for delivery to the PM, early and late endosomes, and in some cells to secretory granules. A COPI-mediated recycling pathway retrieves proteins from the Golgi and ERGIC and returns them to the ER. **(B)** Both the COPII (left) and COPI (right) coats are directed in their assembly by small GTPases of the Arf/Sar1 family. In the COPII coat, Sar1 is activated by its guanine nucleotide exchange factor (GEF), Sec12, which localizes to the ER membrane. Activated Sar1–GTP recruits the Sec23/Sec24 dimer, which corresponds to the "inner coat" layer and provides the cargo-binding function. A heterotetramer of Sec13/Sec31 is subsequently recruited, forming the "outer coat" layer that drives membrane curvature. In the COPII cage formed by Sec13/Sec31, four molecules of Sec31 assemble head-to-head via β-propeller domains to form the "vertex" of the cage (inset). The COPI coat assembles upon activation of Arf1, which is driven by either of the redundant GEFs, Gea1 or Gea2. Arf1 in turn recruits the inner coat complex of Sec21/Sec26/Ret2/Ret3. The COPI outer coat is formed by Sec27/Ret1/Sec28, which assembles in a triskelion structure via interactions of three β-propeller domains of Sec27 (inset) (Adapted from Szul and Sztul, 2011 and Barlowe and Miller, 2013).

1.1.2.2 Motor proteins

After budding, vesicles move towards their final destination by diffusion or by motor-mediated transport along cytoskeletal tracks. Motor proteins that drive vesicles along the cytoskeletal highway include families of cortex myosins, traveling along actin filaments, and cytosolic kinesins and dyneins, traveling on microtubules.

Actin filaments and microtubules are polarized; ideally, microtubules originate from an organizing center (MTOC) near the nucleus and fan out with their plus-ends towards the cell periphery, while actin filaments within the cell cortex are directed with plus-ends towards the cell membrane (Ross et al., 2008) (Fig. 1.3).

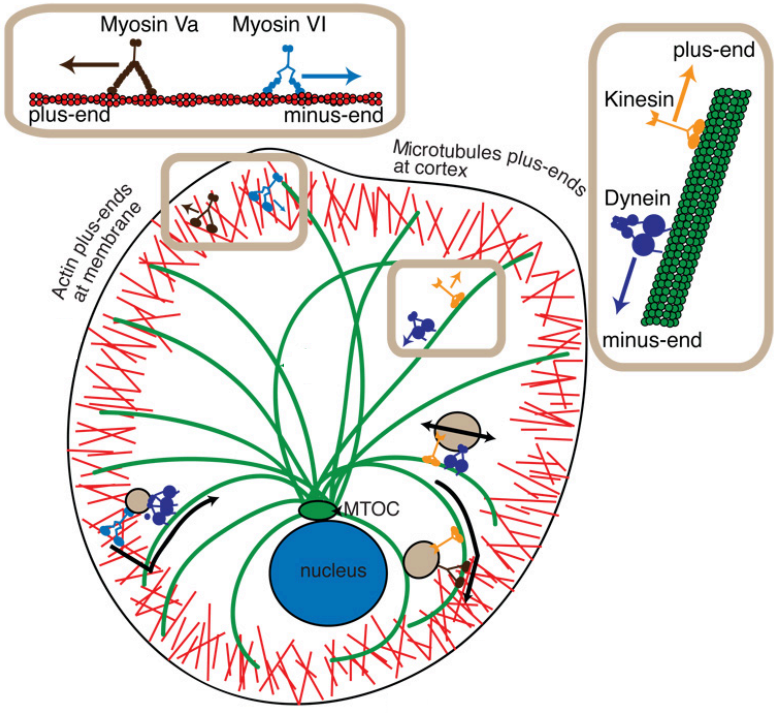


Figure 1.3 Motor proteins and vesicular cargo transport in the cell. Myosin family motors, myosin Va (dark brown) and myosin VI (light blue), walk along actin filaments (red) at the cortex. Microtubule-based motors include the kinesin family motors (orange) and cytoplasmic dynein (violet). Kinesin motors walk to the plus-ends of microtubules (green), which are oriented toward the actin cortex. Dynein motors walk toward the minus-end of the microtubule, which is located at the microtubule-organizing center (MTOC, green) near the cell nucleus (blue) (Modified from Ross et al., 2008).

During exocytosis, transport vesicles, carried by plus-end directed kinesins, are translocated along microtubules toward the cortex. Myosins associate with both microtubule ends and cortical actin filaments (Wu and Bezanilla, 2014), thus vesicles are transferred, at the level of microtubule ends, to anterograde (plus-end directed) myosins V to be delivered to cell membrane. In the reverse direction, during endocytosis, vesicles are transported initially through the actin cortex by the minus-end directed myosins V and then are transferred to retrograde dyneins traveling on microtubules.

1.1.2.3 Tethering factors

Tethering factors are a heterogeneous group of peripherally associated membrane proteins that link transport vesicles with acceptor membrane to ensure correct docking and fusion. Tethers can be grouped into two major classes based on structural similarities: homodimeric coiled-coil tethers and multi subunit tethering complexes (MTCs) (Chia and Gleeson, 2014).

Coiled coil tethers have two globular heads that are joined together by an α-helical coiled-coil rod. These proteins are highly extended and the current model is that they mediate the initial step in vesicle tethering, being able to establish long-range interactions with vesicles (>200 nm). Approximately 20 members are associated with the Golgi (thus termed "golgins"), are regulated by small GTPases of the Rab and Arl families (see section 1.1.3.1), and function in membrane-membrane and membrane-cytoskeleton tethering at the Golgi apparatus; others can be found at endosomes (e.g. early endosome antigen 1 (EEA1)).

MTCs comprise a diverse family of multi subunit proteins. They interact with vesicles over a shorter distance (up to 30 nm) than the coiled-coil tethers and are thought to tether a captured vesicle in close apposition to its acceptor

compartment, namely when it is about to fuse.

MTCs can be further divided into two functional groups: oligomeric complexes that bind to SNAREs (see section 1.1.2.4) and typically act as Rab effectors, comprising Dsl1p, conserved oligomeric Golgi (COG) complex, Golgi-associated retrograde protein (GARP) complex and the exocyst, and oligomeric complexes that function as GEFs for Rab proteins, comprising TRAnsport Protein Particle (TRAPP) complexes and homotypic fusion and protein sorting (HOPS) complex (both a GEF and a Rab effector) (Sztul and Lupashin, 2009).

The TRAPP was first identified in yeast in 1998. The study focused on the protein Bet3, previously identified in a large-scale genetic screen as an essential gene that genetically interacts with ER-to-Golgi SNAREs but is not part of a SNARE complex (Rossi et al., 1995). Using Bet3 as a probe and using a combination of genetic and biochemical studies, the core components of the yeast TRAPP complex were identified. TRAPP components were named with the prefix "trs" (transport subunits) followed by a number corresponding to the molecular weight of each protein: Trs20, Trs23, Trs33, Trs31 and Trs18 (more commonly known as Bet5).

The complex was proposed to participate in ER-to-Golgi trafficking since depletion of Bet3 caused impairment in the arrival of COPII vesicles to the Golgi complex. Furthermore, the trafficking defect could be overcome by over-expression of an ER-to-Golgi specific SNARE suggesting that the TRAPP complex acts upstream of COPII fusion with Golgi membranes (Sacher et al., 1998). Following the identification of an additional four sub-units (Trs65, Trs85, Trs120, and Trs130) and the observation that orthologs are present in mammals (Sacher et al., 2000), a role for TRAPP as a tethering factor that acts during the docking of COPII vesicles with the Golgi was proposed (Barrowman et al., 2000).

Later, Sacher and co-workers postulated the existence of two different TRAPP complexes acting as GEF: TRAPPI, that functions in ER-to-Golgi trafficking by docking COPII vesicles with Golgi membranes and subsequently activating the small GTPase Ypt1, and the larger TRAPPII complex, required for intra-Golgi trafficking, that activates the small GTPases Ypt31/32p (Kim et al., 2016). So far, in addition to TRAPPI and TRAPPII, another two TRAPP complexes have been identified in yeast (TRAPPIII and TRAPPIV) that activate Ypt1 to regulate the assembly of the pre-autophagosomal structure (PAS), the first step of autophagy (Fig. 1.4)

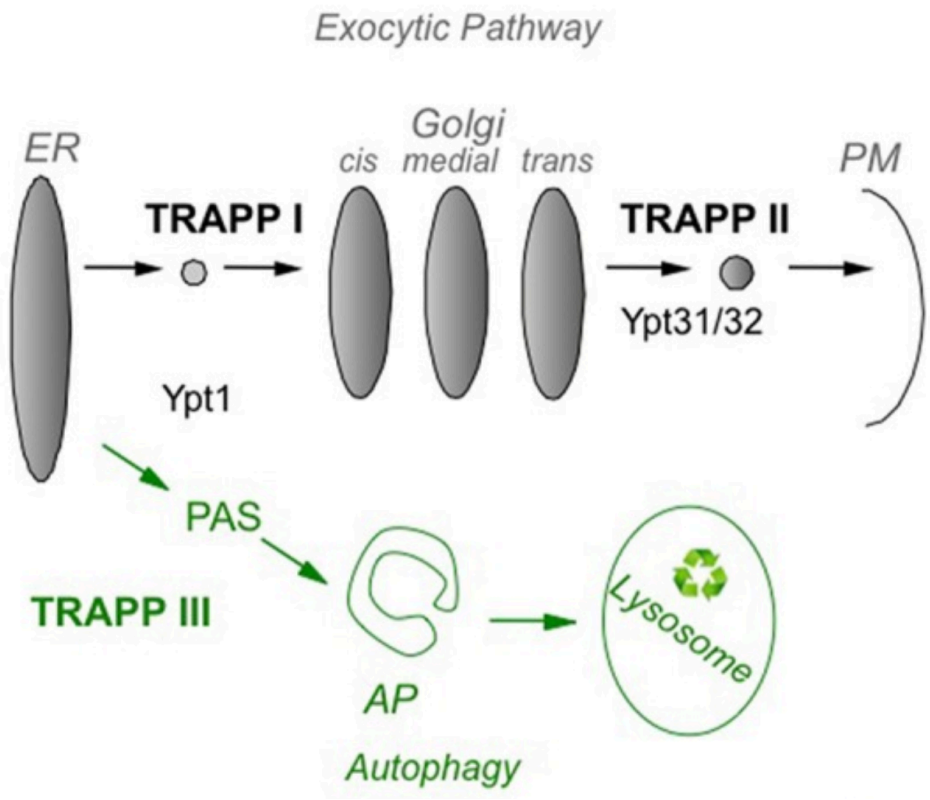


Figure 1.4 The role of TRAPP complexes and their Ypt substrates in yeast intra-cellular trafficking. In the exocytic pathway (top), TRAPP I activates Ypt1 to regulate ER-to-Golgi transport, whereas TRAPP II activates Ypt31/32 to regulate Golgi-to-PM transport. In autophagy (bottom), a cellular recycling pathway (in green), TRAPP III and TRAPP IV activate Ypt1 to regulate the assembly of PAS, the first step of autophagy. PAS is required for the formation of the double-membrane autophagosome (AP), which delivers cargo for degradation in the lysosome (Adapted from Kim et al., 2016).

1.1.2.4 SNAREs

SNAREs (soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptors) are a family of TM proteins, which mediate the final step of vesicle transport, fusion. There are over 30 different SNAREs in mammalian cells, each anchored to the membrane of a particular organelle by a C-terminal TM domain or through lipid modifications such as prenylation (i.e. YKT6) or palmitoylation (i.e. SNAP-23) (Chen et Scheller, 2001; McNew et al., 1997; Vogel et Roche, 1999)

The SNARE hypothesis proposed by (Rothman, 1994), suggests the existence of complementary sets of SNAREs, which assemble in specific complexes to regulate membrane fusion: vesicle (v)-SNAREs, found as a single polypeptide chain on the vesicle membranes, and target (t)-SNAREs, found in complex of two or tree proteins on target membranes. The first SNARE complex, to drive membrane fusion, identified, was that of VAMP1 (Trimble et al., 1988), acting as a v-SNARE on synaptic vesicles, and Syntaxin1 and SNAP-25 (synaptosomal-associated protein of 25 kDa) acting as t-SNAREs on pre-synaptic membranes (Bennett et al., 1992). The X-ray crystal structure of this complex, elucidated by Axel Brunger and Reinhard Jahn (Sutton et al., 1998), revealed a bundle of four parallel alpha helices that formed a pin-like arrangement forcing the two bilayers together as the SNARE complex "zippered up" to result in fusion (Fig. 1.5). This mechanism is called "SNAREpin."

The specific assembly of SNARE complexes is regulated by tethering factors and Sec1/Munc18 family (SM) proteins. Tethering factors gather the formation of fusogenic SNARE complexes and prevent, the assembly of non-fusogenic ones. On the other hand, they influence the stabilization of SNARE complexes, protecting SNAREs from degradation (Wang et al., 2017).

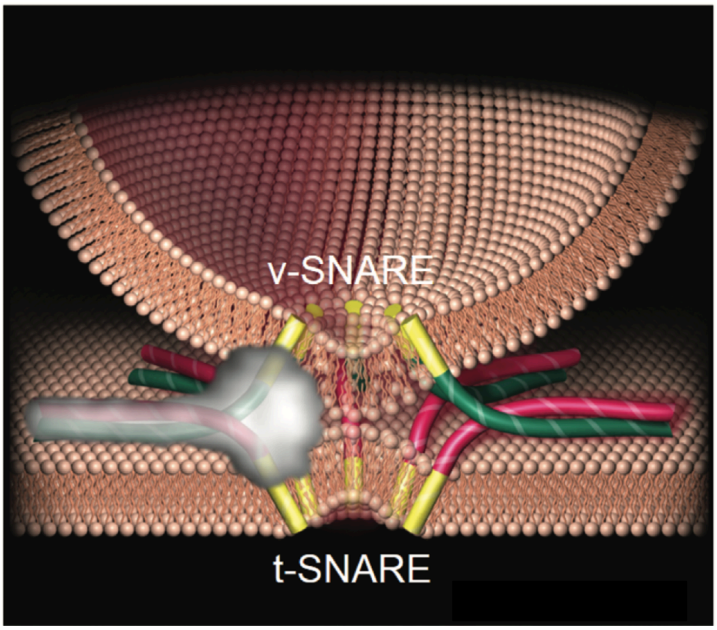


Figure 1.5 The SNARE fusion machinery. v-SNAREs (in green) on a vesicle bind to their cognate t-SNAREs (in red) on the target membrane, forming specific SNAREpins that then fuse the two membranes. For simplicity, the t-SNARE is shown as a single elongated rod, although it is now known to contribute three alpha helices to a four-helix v-t-SNARE bundle. Other proteins regulate the assembly and disassembly of SNAREpins and thus control membrane fusion (Modified from Weber et al., 1998).

MTCs can also cooperate with SM proteins to trigger the assembly and stabilization of SNARE complexes.

SM family comprises 4 members evolutionary conserved: Sec1/Mun18, Sly1, Vps45 and Vps33. Sly1 protect the ER t-SNARE Ufe1p from degradation (Yamaguchi et al., 2002); Vps45 interacts with Tlg2p to prevent its proteosomal degradation (Dulubova et al., 2002); Vps33 (a member of SM proteins and a subunit of the HOPS complex) promotes the assembly of SNARE complexes and inhibits their disassembly by Sec17p/Sec18p, a soluble N-ethylmaleimide-sensitive factor attachment protein (SNAP) and its factor (NSF), which have been proven to disrupt the SNARE complex using energy derived by ATP hydrolysis (Lobingier et al., 2014; Söllner et al., 1993; Xu et al., 2010). Vps33 can also act as a

template by simultaneous binding with Vam3 (t-SNARE) and Nyv1 (v-SNARE) preventing the formation of non-physiological SNARE complexes (Baker et al., 2015).

1.1.2.5 Lipid-transfer proteins

The organelles have a specific lipid composition, which determines the types of TM proteins that will be inserted into their membranes, and the types of peripherally associated proteins that will be recruited to their cytoplasmic surface (e.g. PI effectors).

It is possible to distinguish three major classes of membrane lipids: phospholipids, glycolipids, and sterols. Both phospholipids and glycolipids have hydrophobic acyl chains with different degrees of saturation, and a polar head, consisting of either phosphorylated glycerol or sphingosine, in case of phospholipids, or glycosylated sphingosine, in case of glycolipids. Sterols are a sub-group of steroids, which confer specific properties to the bilayer such as fluidity.

The majority of the organelles are not engaged in the synthesis of these lipids, thus they receive them by either vesicular transport or selective transfer mediated by lipid-transfer proteins (LTPs). In the first case, lipids are included in vesicles, by bulk flow during budding, as part of the membrane. In the second case, instead, the recognition of the lipid cargoes is mediated by a lipid transfer domain, which in general can bind two different lipid molecules, as demonstrated for the Oxysterol-binding protein (OSBP; Mesmin et al., 2013) and strongly predicted for the OSBP-related proteins 5 and 8 (ORP5 and ORP8; Chung et al., 2015). In addition to cholesterol and phosphatidylserine (PS), transported from the ER to the TGN or PM respectively, these proteins also bind PI4P, which is returned to the ER and

hydrolyzed by the PI4P-phosphatase Sac1, providing energy to drive cholesterol or PS against their concentration gradients.

Many LTPs have been identified at the level of membrane contact sites (MCSs), membrane appositions between two organelles, which range from 10 and 15 nm (Holthuis and Levine, 2005; Lev, 2010). The majority of ORP proteins in mammals (including OSBP), for example, have a short-sequence motif containing two phenylalanines in an acidic tract (FFAT motif) that mediates binding to the ER through interaction with the VAMP-associated proteins A and B (VAPA and VAPB), whose yeast ortholog, Ssc2, has been shown to function at ER-PM MCSs (Manford et al., 2012). Mammalian VAPs, conversely, function at ER-Golgi MCSs where they interact with many LTPs, comprising OSBP, the ceramide transfer protein CERT, and the glucosyl-ceramide transporter FAPP2 (Yamaji and Hanada, 2015). Like FAPP2 also OSBP and CERT have PH domains that recognize both PI4P and Arf1 at the TGN (De Matteis and Rega, 2015) (Fig. 1.6). The PH domain of ORP1L, instead, binds to late endosome (LE)-specific phosphoinositides (Rocha et al., 2009) and those of ORP5 and ORP8 bind to the PM pools of PI4P (Chung et al., 2015).

All the phosphoinositides (PIs), PI4P included, will be treated in section 1.1.3.2.

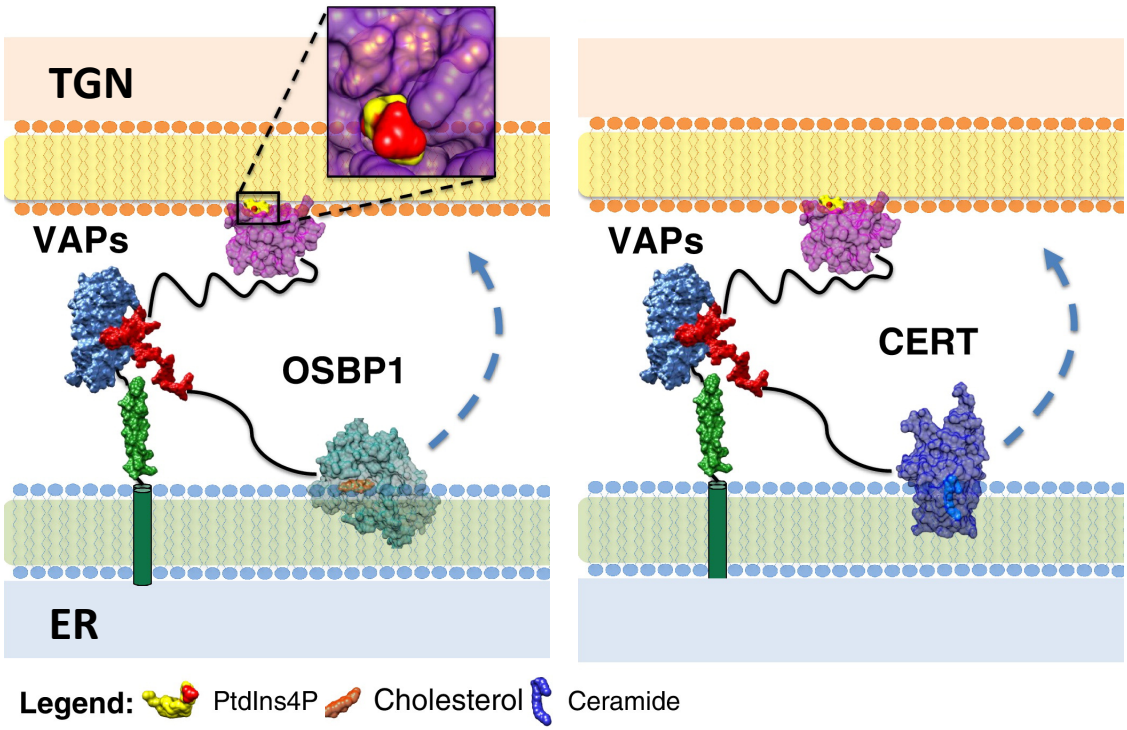


Figure 1.6 Lipid transfer proteins at ER–TGN contact sites. Left: OSBP1 has targeting domains for two membranes, an N-terminal pleckstrin homology (PH) domain (pink) that typically binds PtdIns4P at the TGN and a central FFAT motif (red) that binds the ER proteins VAPs (blue-green). The C-terminal lipid transfer domain (light green) transfers cholesterol from the ER to the TGN (upward pointing arrow). Right: CERT has a domain structure similar to OSBP1, with a PH domain (pink) that binds PtdIns4P at the TGN and a FFAT motif (red) that binds VAPs at the ER. The C-terminal lipid transfer domain (purple) transfers ceramide from the ER to the TGN (upward pointing arrow) (Modified from De Matteis and Rega, 2015).

1.1.3 Regulators of the vesicular transport

The four essential steps of vesicular transport are highly regulated to ensure that vesicles generated from a donor compartment are delivered to their correct acceptor compartment.

1.1.3.1 Monomeric GTPases

Monomeric GTPases, also known as small GTPases, are probably the most important class of membrane trafficking regulators since many steps in vesicular transport depend on a variety of GTP-binding proteins that control both the spatial

and temporal aspects of membrane exchange. These monomeric enzymes contain a GTPase domain for the hydrolysis of GTP, which allows them to act as molecular switches: when GTP is bound, the switch is turned “on” and downstream effectors are activated; hydrolysis of GTP to guanosine diphosphate (GDP) converts the proteins into their inactive conformation and the switch is turned “off ” again (Heider et al., 2010). The conformational switching is subjected to regulation by two classes of proteins: GEFs that activate the proteins by catalyzing the exchange of GDP for GTP and GTPase-activating proteins (GAPs) that inactivate the proteins by triggering the hydrolysis of the bound GTP to GDP.

The family of small GTPases consist of more than 150 members, which based on their sequence homology are divided into several subfamilies such as Sar1/Arf /Arf-like (Arl), and Rab subfamilies. The Arf superfamily comprises regulators of vesicle budding: when in GDP-bound state, these proteins have cytosolic localization, but once switched to their active form, these GTPases expose an amphiphilic helix, which inserts into the cytoplasmic leaflet of the membrane, where they start recruiting coat proteins. Arf proteins are responsible for both COPI and clathrin coat recruitment during retrograde transport in the Golgi and budding from the trans Golgi and the PM; Sar1 is responsible for COPII coat assembly at ER membrane (Molendijk et al., 2004). In addition to the coat-recruitment activity, Arf proteins also regulate lipid-metabolizing enzymes and their product effectors. ARF1, for example, stimulates the synthesis of the phosphatidylinositol 4-phosphate (PtdIns4P or PI4P) by recruiting PtdIns 4-kinase beta (PI4KB) on the Golgi (Godi et al., 1999). Among the PI4P effectors there are the FAPP proteins, which localize to the TGN through coincident detection of both

PI4P and ARF1 mediated by the pleckstrin homology (PH) domain (Godi et al., 2004).

With approximately 70 members identified in humans, Rab proteins are the largest subfamily of such GTPases. Like the Arf GTPases, also Rab proteins cycle between a membrane and the cytosol and their selective distribution on these membranes makes them ideal markers for membrane-enclosed compartments (Zerial and McBride, 2001). Different Rab GTPases assemble specific Rab domains on organelle membrane by recruiting heterogeneous effectors, which facilitate budding, vesicle transport, tethering and fusion. Rab5, for example, assembles on endosomal membranes and recruits tethering proteins, which mediate the capture of CCVs arriving from PM. In addition Rab5 activates a PtdIns 3-kinase, which convert PtdIns to PtdIns 3-phosphate (PI3P), bound by some Rab effectors (Christoforidis et a., 1999; Murray et al., 2002) (see section 1.3.2). On the same membrane, other two Rabs, Rab4 and Rab11, assemble a distinct domain, responsible for the budding of recycling vesicles that return cargoes from the endosome to the PM (see section 1.3.3).

1.1.3.2 PI functions in membrane trafficking

PIs consist of a family of phospholipids that over the last two decades have emerged as key regulatory components of cell membranes in trafficking and cytoskeleton remodelling, despite their low abundance (less than 10% of the total cellular phospholipids) (Di Paolo and De Camilli, 2006; Michell, 2011; Roth, 2004). It has become clear that they serve as constitutive signals that recruit various PI-binding proteins, helping defining organelle identity.

PtdIns is the basic building block for the PIs of eukaryotic cells. It consists of an inositol headgroup (D-myo-inositol 1-phosphate) that is esterified via a

phosphodiester link to 1-stearoyl, 2-arachidonoyl diacylglycerol. PtdIns is synthesised by the enzyme CDP-diacylglycerol-inositol 3- phosphatidyltransferase (known as PtdIns synthase) in the ER, and then delivered either by vesicular transport or lipid-transfer proteins to other membranes, where its subsequent phosphorylation steps occur (Cockcroft and De Matteis, 2001). Only three (at positions D3, D4 and D5) of five free hydroxyl groups in the inositol ring are subjected to reversible phosphorylation, resulting in the generation of seven PI species: 3 monophosphorylated, and 4 polyphosphorylated) (Fig. 1.7).

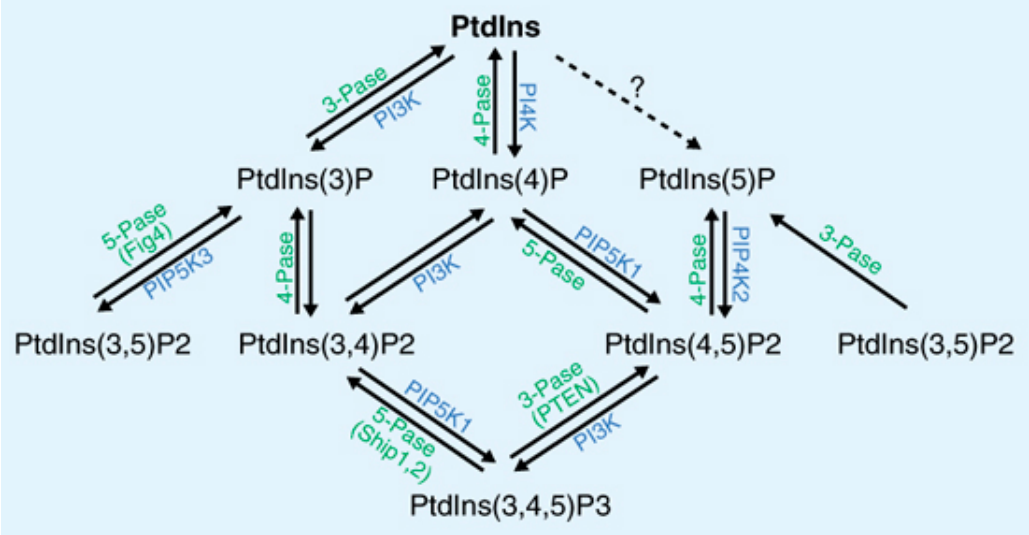


Figure 1.7 PI metabolism. The main pathways of PI synthesis and degradation in mammalian cells are reported with the PI-kinases in blue and PI-phosphatases in green (Adapted from Vicinanza et al., 2008).

PI metabolism is spatially and temporally regulated through controlled recruitment and activation of the controlling kinases and phosphatases, such that distinct PIs can be enriched in specific membrane compartments (De Matteis and Godi, 2004; Vicinanza et al., 2008) (Fig 1.8).

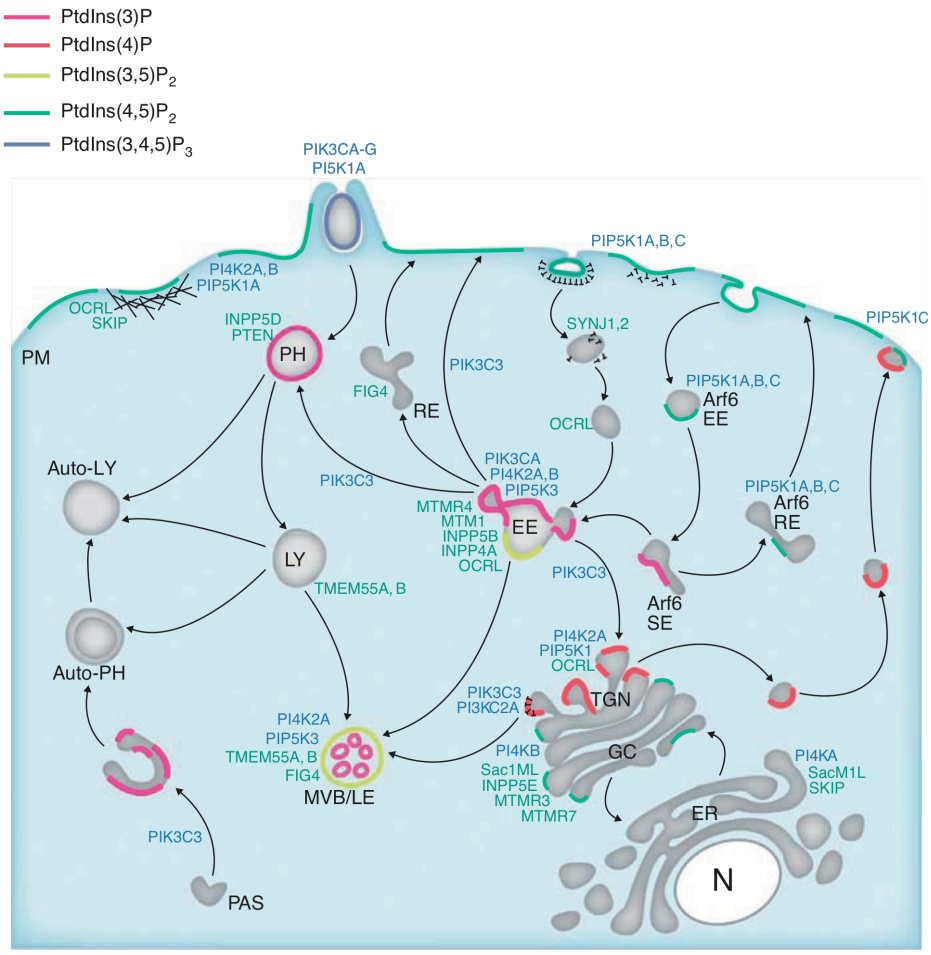


Figure 1.8 Localization and distribution of the PI and the PI-metabolizing enzymes. PI-kinases (blue) and PI-phosphatases (green) are differentially distributed in the cell. PM, plasma membrane; EE, early endosome; SE: sorting endosomes; RE, recycling endosome; LY, lysosome; MVB/LE, multivesicular body/late endosome; PAS, pre-autophagosomal structure; PH, phagosome; TGN, trans-Golgi network; GC, Golgi complex; ER: endoplasmic reticulum; N, nucleus (Adapted from Vicinanza et al., 2008).

In vivo, the PIs concentrate at the cytosolic surface of biological membranes with their lipid portion directly inserted into the lipid bilayer, while the hydrophilic headgroup protrudes into the cytosol. In this way, the headgroup remains accessible for the lipid kinases, phosphatases and phospholipases that regulate the PI turnover, and also for a wide range of cytosolic proteins (including motors, coats, adaptors, GTPase regulators and sorting machineries) that are recruited to membranes in this way.

1.2 Pathways and mechanisms of exocytic membrane trafficking

1.2.1 Polypeptide targeting and translocation

The first step of the secretory pathway is the translocation of synthesized proteins that are destined for the Golgi apparatus, or beyond, into the ER. In yeast, both co-translational and post-translational mechanisms operate to target diverse sets of secretory proteins to the ER. The co-translational process begins with the recognition of an N-terminal signal sequence by the signal recognition particle (SRP), which drives the nascent polypeptide chain into the ER lumen, through SRP receptor (SR) binding (Ogg et al., 1998; Fig. 1.9A). The post-translational insertion of secretory proteins, conversely, depends on cytosolic Hsp70 ATPases, such as Ssa1, that bind the N-terminal signal to maintain the nascent protein in an unfolded translocation competent state until delivery to the Sec63 complex (Chirico et al., 1988; Deshaies et al., 1988; Fig. 1.9B). Both Sec63 and SRP-bound SR complexes require the Sec61 complex to coordinate directed movement of the nascent chain into the ER lumen (Rapoport, 2007). More recently another post-translational translocation pathway to the ER has been described for TM proteins, including the large class of SNARE proteins that drive intracellular membrane fusion events. This pathway, independent of Sec63 and Sec61 complexes, involves a set of genes that produced Golgi-to-ER trafficking deficiencies, thus named GET genes (Schuldiner et al. 2008). Current models for the GET targeting pathway in yeast suggest that an Sgt2–Get4–Get5 sub-complex targets nascent polypeptides to Get3, which in association with an integral membrane Get1/Get2 complex, and in presence of ATP, integrates tailed-anchored proteins into the ER membrane bilayer (Shao and Hegde 2011; Fig. 1.9C).

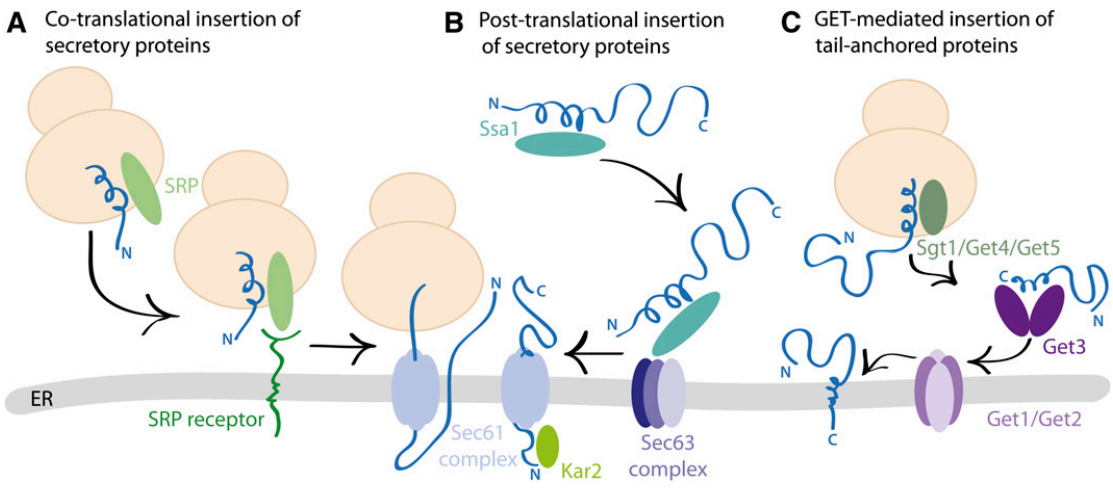


Figure 1.9 Membrane translocation of secretory proteins. Three well-characterized pathways operate to deliver secretory proteins to the ER for membrane translocation. **(A)** Co-translational insertion of secretory proteins. **(B)** Post-translational insertion of secretory proteins. **(C)** GET-mediated insertion of secretory proteins (Adapted from Barlowe and Miller, 2013).

1.2.2 Maturation of secretory protein in the ER

During the translocation through the Sec61 complex, the N-terminal signal sequence of many secretory proteins undergoes the endoproteolytic cleavage of the signal peptidase complex (SPC), composed of four polypeptides termed Spc1, Spc2, Spc3, and Sec11 (Bohni et al., 1988; YaDeau et al., 1991). Coincident with the signal sequence cleavage, an N-linked oligosaccharide of 14 residues, Glc3Man9GlcNAc2 (where Glc=Glucose, Man=Mannose, and GlcNAc=N-acetylglucosamine), is attached to consensus Asn-X-Ser/Thr sites in transiting polypeptides by the oligosaccharyltransferase (OST) enzyme, detected in complex with the Sec61 translocon (Kelleher and Gilmore, 2006). After the glycan transfer, a series of processing steps assist the protein folding in the ER lumen (Helenius and Aebi, 2004): first, glucosidase I and II remove two glucose residues to form the GlcMan9GlcNAc2 structure, recognized by the lectin chaperones, calnexin (CNX)

and calreticulin (CRT); second, further trimming by glucosidase II removes the final glucose residue to result in the release of calnexin/calreticulin interactions; last, the ER-localized mannosidase I removes the terminal mannosyl residue of the Man9GlcNAc2 structure and it is typically the Man8GlcNAc2 form of fully folded glycoproteins that is exported from the ER (Jakob et al., 1998) (Fig. 1.10).

In addition to N-linked glycosylation, some secretory proteins undergo O-mannosylation on Ser/Thr residues by O-mannosyltransferases (Pmts), which is essential in yeast for cell wall integrity as well as normal morphogenesis (Strahl-Bolsinger et al. 1999). Furthermore, almost the 15% of proteins, destined to PM, acquires the addition of a glycosylphosphatidylinositol (GPI) anchor in a multistep process that depends on >20 gene products (Orlean and Menon, 2007).

1.2.3 Transport from the ER

Once properly folded, protein cargoes are packaged into small COPII-coated vesicles that bud from specific sub-domain of the smooth endoplasmic reticulum, known as ER exit sites (ERES), from which they are transported to the Golgi complex.

Cargo molecules can be packaged into COPII carriers using different mechanisms: (I) bulk-flow; (II) cargo capture or (III) the direct binding to Sec24.

(I) Bulk flow export is a passive means of ER secretion where soluble or membrane-associated proteins are stoichiometrically sequestered into COPII vesicles by default. The cargo molecules are simply included as components of the bulk membrane and fluid.

(II) Cargo capture is a process by which, after proper folding, proteins are released from chaperones and bind to specific receptors localized on ERES that mediate their capture and packaging into COPII-nascent carriers (Dancourt and Barlowe,

2010).

(III) TM cargoes can be exported out of the ER by binding directly to the inner COPII coat component Sec24. To accommodate the diversity of cargo that must transit through the ERES, cells express multiple isoforms of the Sec24 adaptor protein, each with up to four non-overlapping cargo recognition sites (Barlowe and Helenius, 2016).

In mammals, once budded, COPII carriers can either fuse with each other (homotypic fusion), forming the ER-Golgi intermediate compartment (ERGIC), absent in yeast, or tether and fuse directly with the intermediate compartment (heterotypic fusion).

The ERGIC is relatively short-lived because it moves quickly along microtubules to the Golgi apparatus, with which it fuses to deliver its contents. While approaching *cis*-Golgi, the ERGIC begins to bud off COPI-coated vesicles that carry back, to the ER, resident proteins either escaped in the bulk flow export or involved in the ER budding reaction and now being recycled. This is possible thank to ER retrieval signals present at the C-terminal ends of these proteins. ER membrane proteins, for example, possess signals, generally consisting of two lysines followed by any two other amino acids (KKXX), that are directly recognized by COPI coats and are thus packaged into COPI-coated vesicles. Soluble ER proteins, instead, possess signals, consisting of a Lys-Asp-Glu-Leu or a similar sequence (KDEL), that are recognized by specialized receptor proteins able to bind the KDEL sequence and package any protein displaying it into COPI-coated vesicles (Stornaiuolo et al., 2003).

A similar retrieval process continues from the Golgi apparatus, after the ERGIC has delivered its cargoes.

1.2.4 Transport through the Golgi apparatus

In the Golgi, secretory proteins may receive other N-linked oligosaccharide modifications as they progress through distinct Golgi compartments.

In the *cis*-Golgi, occurs the trimming of three mannose residues by Golgi mannosidases IA, IB, and IC; in the *medial*-Golgi, the Man5GlcNAc2 chain is further processed by addition of a GlcNAc residue catalyzed by N-acetylglucosaminyltransferase (GlcNAcT) I (Kornfeld and Kornfeld, 1985). The transfer of GlcNAc initiates the synthesis of complex N-glycans, which lose two mannose residues by Golgi mannosidase II and receive another GlcNAc by GlcNAcT-II. The final steps of complex oligosaccharide synthesis occur in the *trans*-Golgi, where the N-glycans are further branched up to six times by the addition of different sugars including galactose (Gal), sialic acid, fucose, GlcNAc, GalNAc and disaccharide units (Stanley et al., 2009) (Fig. 1.10).

It is still uncertain if these molecules move from one Golgi compartment to another by vesicular transport or cisternal maturation (the enzymes present in each individual compartment change over time, while the cargo proteins remain inside the cisterna), although evidences of changes in the protein composition of individual cisternae over time support the maturation model (Losev et al., 2006; Matsuura-Tokita et al., 2006).

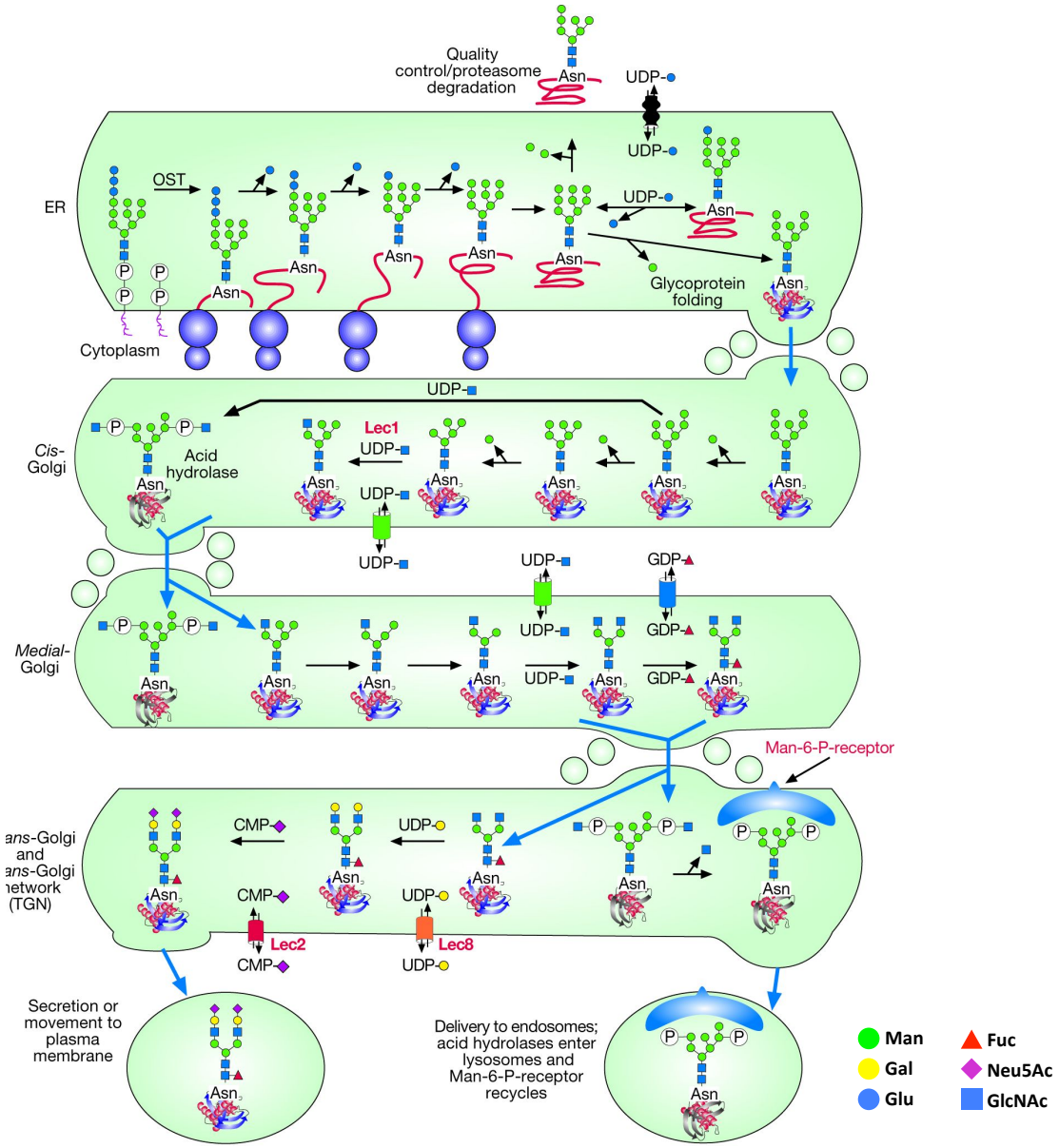


Figure 1.10 Processing and maturation of an N-glycan. Following transfer of the 14-sugar $Glc_3Man_9GlcNAc_2$ glycan to Asn-X-Ser/Thr sites, glucosidases in the ER remove the three glucose residues, and ER mannosidase removes a mannose residue. These reactions are intimately associated with the folding of the glycoprotein assisted by the lectins, calnexin and calreticulin, and they determine whether the glycoprotein continues to the Golgi or is degraded. For most glycoproteins, additional mannose residues are removed in the *cis* compartment of the Golgi until $Man_5GlcNAc_2Asn$ is generated. The action of GlcNAcT-1 on $Man_5GlcNAc_2Asn$ in the *medial*-Golgi initiates the first branch of an N-glycan. α-Mannosidase II removes two outer mannoses and generates the substrate for GlcNAcT-II. The resulting biantennary N-glycan is extended by the addition of fucose, galactose, and sialic acid to generate a complex N-glycan with two branches. Also shown is the special case of lysosomal hydrolases that acquire a GlcNAc-1-P at C-6 of mannose residues on oligomannose N-glycans in the *cis*-Golgi. The *N*-acetylglucosamine is removed in the *trans*-Golgi by a glycosidase, thereby exposing Man-6-P residues that are recognized by a Man-6-P receptor and routed to an acidified, prelysosomal compartment (Adapted from Kornfeld and Kornfeld, 1985).

1.3 Pathways and mechanisms of endocytic membrane trafficking

1.3.1 PM internalization

Endocytic membrane trafficking starts with the cellular internalization of PM components, like receptors and lipids, and extracellular material., like receptor-associated ligands and solute molecules. The advent of new microscopy and biochemical techniques has provided the idea that there are many routes of endocytic uptake into the cell (Doherty and McMahon, 2009), which for simplicity can be divided into two main groups: those that require the GTPase dynamin for newly formed vesicles abscission, and those clathrin- and dynamin-independent (Fig. 1.11).

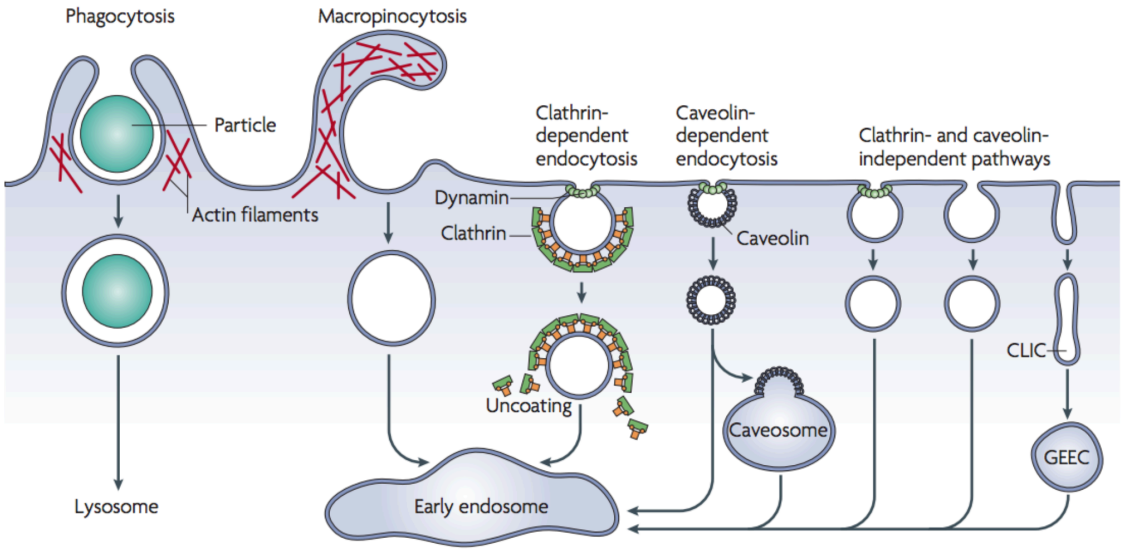


Figure 1.11 Multiple endocytic pathways of PM internalization. Large particles can be taken up by phagocytosis, whereas fluid uptake occurs by macropinocytosis. Both processes appear to be triggered by and are dependent on actin-mediated remodelling of the plasma membrane at a large scale. Compared with the other endocytic pathways, the size of the vesicles formed by phagocytosis and macropinocytosis is much larger. Most internalized cargoes are delivered to the early endosome via vesicular (clathrin- or caveolin-coated vesicles) or tubular intermediates (known as clathrin- and dynamin-independent carriers (CLICs)) that are derived from the plasma membrane. Some pathways may first traffic to intermediate compartments, such as the caveosome or glycosyl phosphatidylinositol-anchored protein enriched early endosomal compartments (GEEC), en route to the early endosome (Adapted from Mayor and Pagano, 2007).

The best-studied pathway of receptor internalization is that mediated by clathrin-coated pits (CCPs; clathrin-dependent endocytosis, CDE; Fig. 1.11).

Small areas of the PM are covered on the cytoplasmic surface with clathrin triskelions that start to assemble in the polyhedral clathrin structure (Fotin et al., 2004). The cytoplasmic domains of PM receptors contain sorting sequences that are specifically recognized by adaptor proteins, which include the clathrin coat adaptors AP2, Dab2, ARH and Epsin (Traub, 2009). These proteins recruit and package the receptors into CCPs that invaginate (inwards) with the help of several accessory proteins, and then pinch off to form a CCV in a process that requires dynamin. The receptors for ironbound transferrin (TfR) and low-density lipoprotrein (LDLR) are classic examples of cargo proteins in CDE.

Many cell-surface TM proteins lack cytoplasmic sequences for their recruitment into CCVs, and they have been shown to be internalized by clathrin-independent endocytosis (CIE). Examples of these receptors are the major histocompatibility complex (MHC) class I proteins, β-integrins, the GPI-anchored protein CD59, and the glucose transporter GLUT1, which have been reported to enter the cell via dynamin- and clathrin-independent endocytosis associated with the Arf6 GTPase (Naslavsky et al., 2004b), and the interleukin-2 receptor (IL-2R), which is internalized in a dynamin-, endophilin-, and RhoA-dependent manner (Boucrot et al., 2015).

Additional non-clathrin-mediated pathways include caveolin and flotillin-dependent endocytosis and internalization of fluid or large extracellular particles, like macropinocytosis and phagocytosis (Fig. 1.11).

Caveolins are a family of proteins that coat PM invagination of 55-60 nm diameter, forms higher-order oligomers and bind cholesterol and fatty acids, which stabilize

oligomer formation (Parton et al., 2006). These properties may be important in ordering local lipids into invagination-competent compositions (called caveolae) and thus in modulating membrane curvature. Caveolar endocytosis is dynamin-dependent and is used by glycosphingolipids and by some viruses for cell internalization (Shin and Abraham, 2001). Despite these observations, the precise role of caveolins, the specific lipid composition of caveolae and the specific functional relevance of other caveolae-associated proteins remain to be firmly established.

The flotillins are a family of coat proteins found in PM-domains distinct from caveolae. Flotillin-2, together with dynamin, has been reported to be necessary for the GPI-linked proteins and cholera toxin B (CTxB) uptake (Ait-Slimane et al., 2009), while flotillin-1 and dynamin-2 have been shown to be required for the internalization of proteoglycans and their ligands (Payne et al., 2007).

Macropinocytosis is dynamin and GTPase Rac1-dependent, with cholesterol required for the recruitment of activated Rac1 (Kerr and Teasdale, 2009). Macropinocytosis derived compartments are called macropinosomes; these are large endocytic compartments involved in fluid-phase internalization and that form upon stimulation of grow factors (e.g. Epidermal growth factor (EGF) and Platelet-derived growth factor (PDGF)), when the tip of membrane ruffles can fuse back with the PM (Orth et al., 2006).

Phagocytosis is the uptake of relatively large particles (300 nm to several μm in diameter) that occurs most robustly in specialized cells, such as macrophages (Niedergang and Chavrier, 2004). In macrophages, large diameter (up to 2 μm) tubules bud into the cytoplasm progressively invaginating opsonized and particulate material intended for internalization.

Lastly, a CIE pathway that is regulated by GTPases Cdc42 and Arf1, and actin, has been reported for fluid-phase markers, GPI-anchored proteins and bacterial toxin (such as CTxB) endocytosis (Kumari and Mayor, 2008). This occur via tubular invaginations at the PM, which are known as clathrin and dynamin-independent carriers (CLICs) and which subsequently undergo fusion, to generate distinct early endosomal compartments: the GPI-enriched early endosomal compartments (GEECs).

1.3.2 Sorting of the endocytosed cargoes

Regardless of the mode of entry, endocytosed cargoes are delivered to a common, Rab5 containing, sorting station, from where the cargo destined for recycling, degradation or retrograde transport are separated out. This common sorting station is known as the early endosomes (EEs) or the sorting endosomes (SEs) (Jovic et al., 2010; Mayor et al., 1993). The small GTPase Rab5, its effectors (EEA1) and PtdIns 3-kinase type III (PI3KIII) and its product PtdIns3P mark the EE and are required for its function. Morphological and biochemical properties of EEs support their complex functions. The lumen of the EEs is mildly acidic (PH~6.3-6.8), thereby facilitating conformational changes that allow for the dissociation of ligands from their receptors. The EEs have a highly dynamic structure, containing a vacuolar part, with large vesicles (~400 nm diameter) and membrane invaginations, from which a reticulum of multi-branching thin tubules (~60 nm diameter) emerges (Bonifacino and Rojas, 2006; Gruenberg, 2001) (Fig. 1.12). These morphologically distinct EE sub-domains are believed to be functionally important, such that cargoes to be recycled cluster within primarily tubular membranes, whereas cargoes destined for lysosomes remain in the EE vacuole (Mellman, 1996). The tubular endosomal network is further sub-

compartmentalized into domains that can lead receptors to enter directly into fast recycling pathways, or to be transferred to a later recycling compartment (Fig. 1.12).

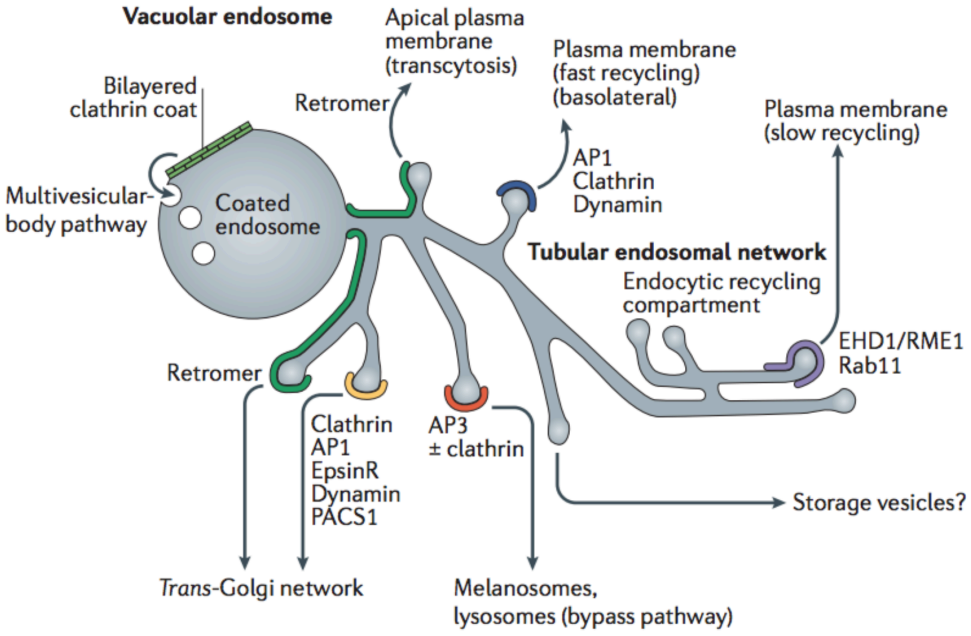


Figure 1.12 Ultrastructure of the sorting endosomes. Once transported to the EEs, cargoes are segregated, by a series of soring events, into separate trafficking itineraries, having as final destination the TGN, PM and lysosome. The EEs morphology consists of a vacuolar region connected to a network of tubules. The vacuolar domain is coated by flat clathrin that has a role in recruiting the ESCRT machinery for targeting proteins to MVBs pathway. The tubular region (tubular endosomal network, TEN) has many domains that contain specific devices (highlighted by different colors) for carrier formation and cargo recycling to TGN or PM (Adapted from Bonifacino and Rojas, 2006).

The critical role of the EE necessitates that any sorting that occurs here is precisely and highly controlled by key components, such as proteins and lipids. Numerous sorting devices include some effectors and regulators of vesicle transport discussed in section 1.1.2 and 1.1.3, for examples GTPase (Rab and Arf) and their effectors, the PI, clathrin and its adaptors, and motor proteins; new ones will be presented in the following sections (1.3.3-1.3.5).

1.3.3 Recycling pathways

Cargoes destined to return to PM enter the recycling pathways. It is the case of TfR recycling, whose kinetic studies have confirmed the existence of two routes with different kinetics: the 'fast' recycling pathway (t1/2=5 min) and the 'slow' recycling pathway (t1/2=15-30 min) (Hopkins and Trowbridge, 1983).

TfR undergoing fast recycling is quickly sorted away from Rab5-containing micro-domains into a Rab4 micro-domain (Sheff et al., 1999; van der Sluijs et al., 1992); alternatively, TfR undergoing slow recycling is delivered to a Rab4/Rab11-containing endocytic recycling compartment (ERC) (Sonnichsen et al., 2000) (Fig. 1.13). The existence of Rab4-dependent fast recycling has been documented also for the transport of glycosphingolipids (Choudhury et al., 2004).

Although Rab4 and Rab5 share multiple effectors, the organization of the recycling domains at EEs appears to be only mediated by Rab4, as immunodepletion of Rab5 itself has no effects (Pagano et al., 2004). Indeed, a Rab4 endocytic regulatory role is achieved by its interactions with the multivalent Rab4/Rab5 effectors, Rabenosyn-5 and the Rabaptin-5/Rabex-5 complex (de Renzis et al., 2002; Mattera and Bonifacino, 2008; Vitale et al., 1998).

Rabaptin-5 functions as a linker between Rab4 and the γ1-adaptin subunit of the clathrin adaptor AP1 (Deneka et al., 2003). This interaction potentially has a regulatory role in preventing the association of clathrin with AP1 γ1-adaptin, which would result in inhibition of coat formation and of budding of EE membranes. Rabenosyn-5, instead, facilitates Rab4 in the delivery of recycling cargoes to the ERC, acting as a linker between Rab4 and the Eps15-homology domain-containing protein EHD1, an endocytic regulator that is found primarily on the tubule-vesicular ERC membranes (Naslavsky et al., 2004a). Depletion of

Rabenosyn-5 leads to the accumulation of TfR in EEs (the EEA1 and Rab5-containing compartment), while depletion of EHD1 results in cargoes accumulation at the ERC.

Studies in mammalian cells have added another Rab, Rab35, to the fast recycling pathway of the TfR (Kouranti et al., 2006); Rab35 is also associated with tubular recycling endosomes, emanating from the ERC, that contain Arf6 and EHD1 and carry specific CIE cargoes back to the PM (Caplan et al., 2002; Walseng et al., 2008). However, a lot of evidence exists for requirements for Rabs, their regulators and effectors, and motors and adaptors in the transport of endocytosed cargo from the EEs to the Rab11-associated ERC, and/or in regulating the slow recycling pathway (Fig. 1.13).

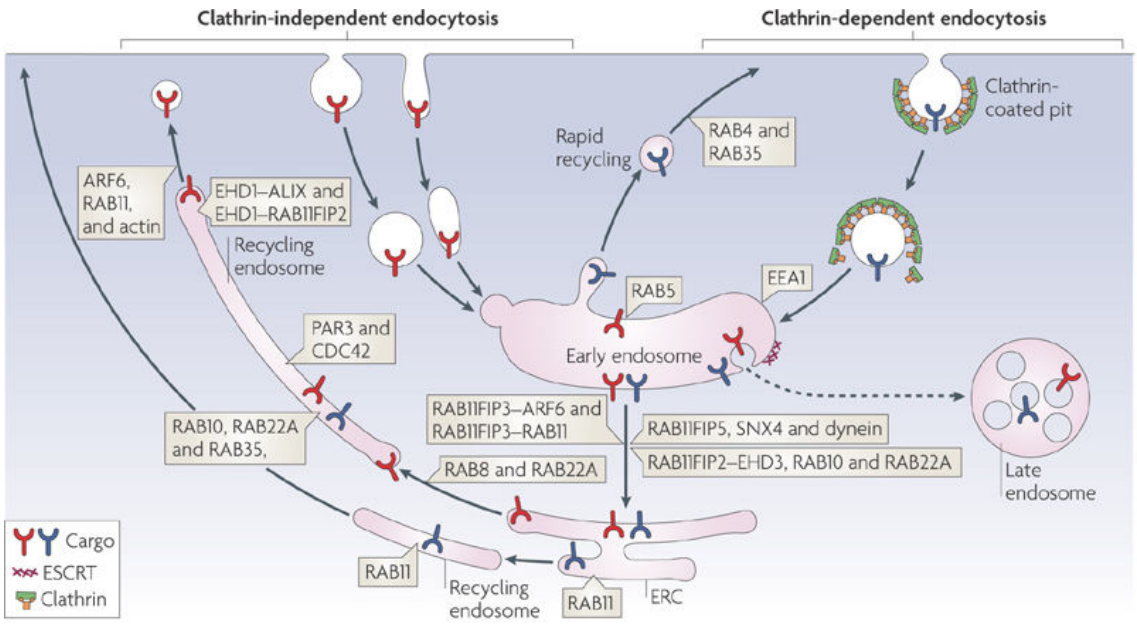


Figure 1.13 Regulation of endocytic recycling pathways. Post-endocytic itineraries of receptors entering the cells by both clathrin-dependent (blue cargo) and clathrin-independent (red cargo) endocytosis are shown. Subsequent routing of cargo to the PM, directly or through the endocytic recycling compartment (ERC), requires several machineries, including GTPases (Rabs and Arfs) and their effectors (i.e. Rab11FIPs) (Adapted from Grant and Donaldson, 2009).

Rab8 co-localize with Arf6 on tubular endosomes that contain CIE cargoes, and it promotes recycling and cortical-actin-driven PM-protusions (Hattula et al., 2006); Rab22a depletion inhibits slow TfR recycling, while Rab10 specifically regulate recycling pathways in polarized epithelial cells (Babbey et al., 2006; Magadan et al., 2006). Rab11-family-interacting proteins (RAB11FIPs) have been identified complexed with proteins that regulate the slow recycling pathway (Inoue et al., 2008; Naslavsky et al., 2006; Schonteich et al., 2008).

1.3.4 Degradative pathway

Cargoes destined to degradation enter the route to lysosome and a well-studied mechanism, in this regard, is that of the EGF-bound receptor (EGFR).

EGFR, lysine mono-ubiquitylated on its cytosolic tail, is concentrated in a sub-domain of the vacuolar EE that has a bilayer clathrin coat, where the hepatocyte growth factor-regulated tyrosine kinase substrate (Hrs) also localizes (Raiborg et al., 2002). Through selective binding to clathrin and recognition of ubiquitylated cargoes, Hrs drives the formation of a unique, Rab7-conteining, sorting micro-domain that actively concentrates, and sorts, cargoes into intraluminal vesicles and, away from the tubular endosomal network and the ERC, actively recycles membranes. Additional ubiquitin-binding proteins, like Eps15 and signal transducing adaptor molecule 2 (STAM2), are part of the complex (Bache et al., 2003). Hrs recruits Tsg101, a component of the endosomal sorting complex required for transport (ESCRT) I, which subsequently triggers ESCRT-II and ESCRT-III assembly, leading to the onset of membrane invagination and formation of multivesicular bodies (MVBs) (Babst et al., 2002a; Babst et al., 2002b; Bonifacino and Hurely, 2008).

The ESCRT complexes not only play a fundamental role in the MVB biogenesis, but also in other cellular processes, such as cellular abscission and viral budding. The unique structure of each complex, enable distinct biochemical function: ESCRT-I acts as a bridge between the functional sorting complex, consisting of Hrs and other ubiquitin-binding proteins (also called ESCRT-0), and ESCRT-II complex, and together with ESCRT-II participates in bud formation; ESCRT-II is probably the most important of the super-complex ESCRT-I/ESCRT-II in MVB biogenesis, since its overexpression can rescue deletions of ESCRT-I functions, but not vice versa; ESCRT-III associates with ESCRT-II and pinches the cargo containing vesicle closed (Hurley, 2010).

Ultimately, fusion of mature MVBs (also known as LEs) with the lysosomal vesicles carrying proteolytic enzymes results in degradation of EGFR and other sorted receptors.

1.3.5 Retrograde pathway

While the formation of endosomal tubular structures facilitates efficient recycling of receptors back to the plasma membrane, retromer-mediated tubulation is essential for retrograde transport of cargoes (like mannose 6-phosphate receptor (MPR)) from EE to TGN (Bonifacino and Hurley, 2008). The mammalian retromer comprises two distinct sub-complexes: the SNX sub-complex, a dimer composed of a combination of SNX1, SNX2, SNX5 and SNX6, and the vacuolar protein sorting (Vps) sub-complex, a heterotrimer composed of Vps26, Vps29 and Vps35 (Collins, 2008; Haft et al., 2000; Hierro et al., 2007).

SNX proteins contain a PtdIns3P-binding Phox (PX) domain, which is essential for association with EE-membranes (Burda et al., 2002; Carlton et al., 2005a; Cozier et al., 2002), as well as a curvature-sensing Bin-Amphiphysin-Rvs (BAR) domain,

which is essential for retromer-regulated formation of tubular membrane structures (Carlton et al., 2004; van Weering et al., 2010). On the other hand, Vps sub-complex, and in particular Vps35, functions to recognize sorting signals within the cytosolic tails of retrograde cargoes, such as MPR (Arighi et al., 2004; Nothwehr et al., 2000; Seaman, 2007).

Depletion of retromer subunits by siRNAs prevents MPR retrieval., leading to re-routing of MPR to lysosomes (Carlton et al., 2005b; Rojas et al., 2007; Seaman, 2004).

AIMS

It is clear that the intracellular transport system is remarkably complex as it encompasses a wide variety of pathways for the transport of cargoes (proteins and other macromolecules) to different cellular compartments. In addition, cells have evolved multiple strategies that use different molecules to transport cargo through the same trafficking pathway. The fundamental unit of this eclectic system is the membrane trafficking module (**MTM**) made up of three essential components: cargoes, effectors and regulators.

Over the last two decades genome-wide and high-throughput siRNA screens have identified many of these MTMs (De Matteis et al., 2013; Liberali et al., 2014; Simpson et al., 2012); however, a note of caution has to be used in interpreting the results of these studies as they suffer from the major limitation of having been performed in immortalized cell lines, such as HeLa and A431 cells, often following artificial cargoes. Thus, the main aims of my project were:

- To define a set of MTMs in more physiological contexts;
- To understand their specific function and transcriptional regulation.

Moreover, membrane trafficking is often regarded as a constitutive process with a high degree of functional redundancy but this view is challenged by the fact that mutations of single trafficking genes with ubiquitous expression give rise to tissue-specific human diseases (De Matteis and Luini, 2011). According to this evidence, another important aim of the project was:

- To investigate the existence of specific MTMs containing disease genes in those tissues that would exhibit the relevant phenotype.

Here, a systems biology approach is proposed to meet the objectives listed above.

1.4 State of the art in module detection, validation and characterization

1.4.1 RNA Sequencing for Gene Expression Profiling

1.4.1.1 Gene expression profiling methods

Gene co-expression networks represent a valid systems biology tool for detecting functional gene modules (Ruan et al., 2010; Ruprecht et al., 2017). The input data for constructing a gene co-expression network is a gene-expression matrix that describes the pattern of genes actively transcribed under specific conditions (e.g. entire tissues or cell lines).

Techniques used to measure gene-expression profiles include "direct" methods like DNA microarrays (Taub et al., 1983) and sequence-based techniques like serial analysis of gene expression (SAGE; Velculescu, 1995). More recently, direct sequencing of transcripts by high-throughput sequencing technologies (RNA-Seq) has become an additional alternative to microarrays and has replaced SAGE (Wang et al., 2009). Like SAGE, RNA-Seq does not depend on genome annotation for prior probe selection and avoids biases introduced during hybridization of microarrays; however, compared to SAGE, it provides a far more precise measurement of the levels of gene transcripts and their isoforms.

Current RNA-Seq methods rely on complementary DNA (cDNA) synthesis from a population of RNA (total or fractionated, such as poly(A)+). This collection of cDNA (referred to as a library) is then amplified and sequenced in a high-throughput manner to produce millions of short sequence reads typically 30-400 bp long, which correspond to individual cDNA fragments. Following sequencing, the resulting reads are either aligned to a reference genome or transcriptome, or assembled de novo. This produces a genome-scale transcription map consisting of

both the transcriptional structure and the level of expression for each gene (Wang et al., 2009).

1.4.1.2 Pre-processing of raw RNA-Seq data

A significant factor to consider when manipulating raw RNA-Seq data is the sequencing depth or library size, which is the number of sequenced reads for a given sample.

Often, a small fraction of highly expressed genes account for large proportions of the sequenced reads, removing power to detect the expression of lowly expressed genes. In addition, multiple sequencing runs might result in differences in depth among samples. These issues leave the experimenter with some choices to make regarding the data analysis:

1. Which read counts to include in the analysis and which to discard;
2. Which normalization methods to use to mitigate biases across samples (Lin et al., 2016).

In general, the choice of a cutoff of the counts per million (CPM) of mapped reads is arbitrary but it is required to remove the extremely noisy data in the expression estimates; conversely, a number of standardized normalization procedures have been developed over recent years to facilitate accurate comparisons between sample groups. Initial methods of normalization modeled the read counts associated to each gene as a Poisson distribution (Marioni et al., 2008). It is important to note that a Poisson distribution models both mean and variance using a unique parameter. However, in biology the variance of gene expression across multiple biological replicates is larger than its mean expression values, a problem known as over-dispersion (Rapaport et al., 2013), which is why other normalization methods are in use today.

These methods include the trimmed mean of M-values (TMM) (Robinson and Oshlack, 2010) and the DESeq normalization (Anders and Huber, 2010). Both TMM and DESeq models assume that the number of reads in sample *j* that are assigned to gene i can be modeled by a negative binomial (NB) distribution:

$$K_{ij} \sim NB(\mu_{ij}, \phi), \quad (1.1)$$

with mean μ_{ij} and variance $\phi = \sigma_{ij}^2$.

The relation between the variance ϕ and the mean μ is generally defined as:

$$\phi = \mu + \alpha\mu^2, \quad (1.2)$$

where α is the dispersion factor.

Estimation of this factor is one of the fundamental differences between the TMM and the DESeq normalization methods. TMM estimates α as a weighted combination of two components, a gene-specific dispersion effect and a common dispersion effect calculated from all genes. DESeq, on the other hand, breaks the variance estimate into a combination of the Poisson estimate (that is the mean expression of the gene) and a second term that models the biological expression variability (Rapaport et al., 2013).

1.4.2 Weighted correlation network analysis

A good number of methods have been developed for constructing and analyzing gene co-expression networks; among these, weighted correlation network analysis (WGCNA; Langfelder and Horvath, 2008a) is by far the most used method for studying biological networks. A flowchart of the main steps of WGCNA is presented in Figure 1.14.

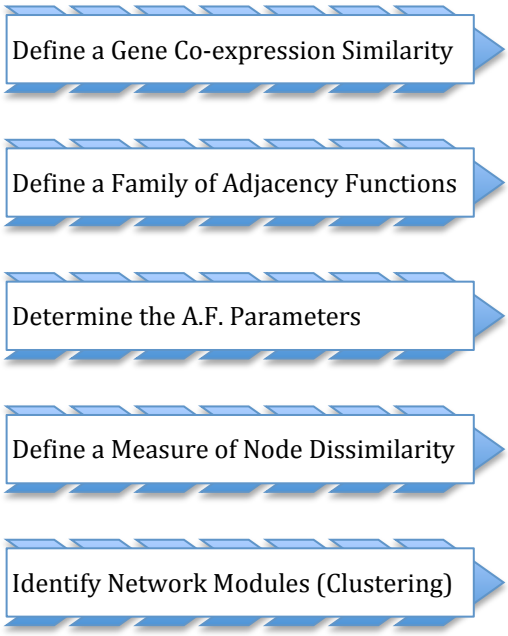


Figure 1.14 Flowchart of gene co-expression network analysis.

1.4.2.1 Definition of a gene co-expression similarity

In the first step, a similarity score between the m genes in the expression matrix is calculated. The result, called similarity matrix S, has $m \times m$ dimensions and in general stores the pairwise correlation coefficients between all the genes.

The Pearson correlation coefficient is one of the measures of linear correlation between two variables X and Y, giving a value between -1 and +1 inclusive, where 1 is total positive correlation, 0 is no correlation, and -1 is total negative correlation. It is the covariance of the two variables divided by the product of their standard deviations:

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \, \sigma_Y}. \qquad (1.3)$$

Another measure is the Spearman correlation coefficient; it is defined as the Pearson correlation coefficient between the ranked variables:

$$\rho_{rgX,rgY} = \frac{\sigma_{rgX,rgY}}{\sigma_{rgX}\,\sigma_{rgY}}, \quad (1.4)$$

where rgX and rgY are the ranked variables of X and Y, respectively.

1.4.2.2 Definition of an adjacency function

In the second step, either a hard or a soft threshold is applied to the similarity matrix to determine the biological meaningfulness of the connections; this choice determines whether the resulting network will be unweighted (hard-threshold) or weighted (soft-threshold).

A currently used adjacency function that implements hard-thresholding is the signum function, defined as:

$$a_{ij} = \begin{cases} 1 \; if \; s_{ij} \geq \tau \\ 0 \; if \; s_{ij} < \tau \end{cases}, \quad (1.5)$$

where s_{ij} is the similarity score between the genes i and j, and τ is the threshold parameter.

An alternative solution, proposed by several authors, is to threshold the significance level (p-value) of the correlation, which is typically estimated by using the Fisher transformation (Davidson et al., 2001) or a permutation test procedure (Butte and Kohane, 2000; Carter et al., 2004), rather than the correlation itself.

However, the major limitation of hard-thresholding functions is that they may lead to a loss of information. In fact, interactions are never black or white (1s or 0s). For this reason, it is often preferred to use a soft-thresholding function, whose intent is not to delete connections but to emphasize stronger associations (larger correlation coefficients).

A widely used 'soft' adjacency function is the power function, defined as:

$$a_{ij} = |s_{ij}|^{\beta}, \quad (1.6)$$

where β is the power parameter.

1.4.2.3 Definition of node dissimilarity and module detection

An important aim of co-expression network analysis is to detect subsets of nodes (modules) that are tightly connected to each other. The operation of identifying similar groups of nodes in a data set is called clustering. Many of the clustering algorithms build models based on dissimilarity measures (Shirkhorshidi et al., 2015), and, in the context of biological networks, one of the most used is the average-linkage hierarchical clustering.

In average-linkage hierarchical clustering, each node is first assigned to its own cluster; then, at each step of the algorithm, the distance between the clusters, defined as the average distance between each point in one cluster to every point in the other cluster, is calculated, and the two most similar clusters are joined together. The final result is a dendrogram whose branches are cut, in a process referred to as tree cutting, branch cutting, or branch pruning, in order to identify the modules.

Distance measures have significant influence on clustering results (Jaskowiak et al., 2014); typically, WGCNA uses the topological overlap dissimilarity measure (Ravasz et al., 2002), defined as:

$$1 - TOM_{ij} = 1 - \frac{\sum_{u \neq i,j} a_{iu} a_{ju} + a_{ij}}{\min\{\sum_{u \neq i} a_{iu}, \sum_{u \neq j} a_{ju}\} + 1 - a_{ij}}, \quad (1.7)$$

where TOM_{ij} is a measure of interconnectedness of genes i and j, and a_{iu}, a_{ju}, and a_{ij}, are the adjacency scores between the corresponding pair of genes.

1.4.3 Validation of predictive models

Validation is the process of assessing the quality of an inferred model with available knowledge. For quantitative validation, it is necessary to employ a scoring methodology that evaluates the model with respect to (a) information

already used to generate the model (internal validation) and (b) information independent from the information used to construct the model (external validation).

The assessment of inferred gene networks generally requires benchmark data sets for which the underlying network is known. Benchmarking involves counting the number of correctly predicted links (true positives, TP), the number of incorrectly predicted links (false positives, FP), the number of correctly identified non-links (true negatives, TN) and the number of true links missed by the inferred network (false negatives, FN). Performance of the model can then be summarized by the receiver operating characteristic (ROC) curve, whose point coordinates are the true positive rate (TPR= TP/TP + FN) and the false positive rate (FPR=FP/FP +TN), respectively (Penfold and Wild, 2011).

In many clustering applications, instead, one may be interested in determining whether the inferred cluster(s) can also be found in other data sets. Statistics for measuring module or cluster preservation belong to two different categories: network-independent and –dependent preservation statistics (Langfelder et al., 2011). Network-independent statistics do not require that a network has been defined, since they aim to evaluate similarity between the cluster assignments of both the reference and the test sets. Network-dependent statistics, by contrast, aim to evaluate that some properties of the inferred modules are maintained in the test networks, i.e. cohesiveness (density preservation statistics) and topology (connectivity preservation statistics).

1.4.4 Characterization of co-expressed gene sets

Analysis of co-expressed gene sets typically involves testing for enrichment of different annotations or 'properties' such as biological processes, pathways, and transcription factor binding sites. It is a common task in bioinformatics to use

enrichment methods, such as a hypergeometric test or gene set enrichment analysis (GSEA) (Subramanian et al., 2005), to characterize genes associated with the co-expressed gene set, and transfer their biological annotations to the related set, something known as guilt-by-association (GBA) heuristic (Wolfe et al., 2005).

In recent years, several GBA methods for the characterization of co-expressed genes have been proposed, many of them involving random walks with restart (Blatti and Sinha, 2016; Hofree et al., 2013; Hou and Ma, 2014; Ivan and Grolmusz, 2011; Li et al., 2017).

A random walk (Lovasz, 1993) is a stochastic process that starts from a given vertex i of a graph G=(V, E), with V the set of vertices and E the set of edges, and then selects one of its neighbors j, uniformly at random, to visit. Running the algorithm for a sufficiently long time, the probability of visiting a particular node j is denoted as the steady state probability of node j, $u(j)$. The steady state probability of a random walk on G is defined as:

$$u_{k+1} = M_G u_k, \tag{1.8}$$

where M_G is the adjacency matrix of G, column normalized for the vertex degrees.

Compared to a random walk, a random walk with restart has an additional probability c of jumping back to one of the nodes of interest (seeds). In this case, the steady state probability of a random walk with restart on G is defined as:

$$u_{k+1} = (1 - c)M_G u_k + cq, \tag{1.9}$$

where q is the vector of seed restarting probabilities.

Genes and properties are prioritized for their relatedness to the seed genes based on this steady state probability; the larger the probability value is, the higher the rank position.

Chapter 2

Material and methods

2.1 Employed datasets

2.1.1 Traffickome list

We have manually curated a list of 1261 genes that cover all the genes described as part of membrane trafficking machineries at the ER, Golgi, Endosomes and Lysosomes. We called this list "traffickome", with reference to the portion of the genome encoding for the proteins involved in the vesicular trafficking. The classes covered by the traffickome, and the percentage of genes falling in each class, are provided in Figure 2.1.

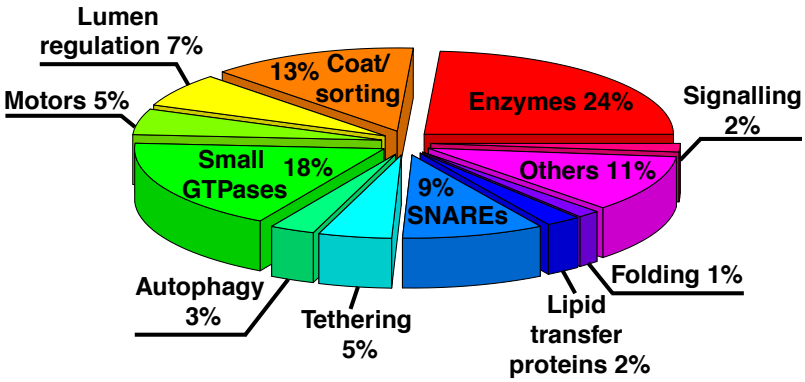


Figure 2.1 Pie chart showing the classes covered by the traffickome.

2.1.2 GTEx V4 release in numbers

We analyzed the Genotype-Tissue Expression (GTEx; GTEx Consortium, 2013) V4 release, which comprises RNA-Seq. data for 2.921 samples, representing 42 body sites (28 solid organ tissues, 11 brain sub-regions, whole blood, and two cell lines: Epstein-Barr virus–transformed lymphocytes (LCL) and cultured fibroblasts from skin; Fig. 2.2), from 209 donors. Of these, 135 (65%) were male, 74 (35%) were

female, 74% White, 35% Black or African American, and 1% Asian. Their ages range from 21 to 70 years, with an average age of 50 years. Death was recorded as due to ventilator death (64%), expected death after illness (10%), unexpected death after illness (5%), fast death of natural causes (10%), violent and fast death (5%), other causes (6%).

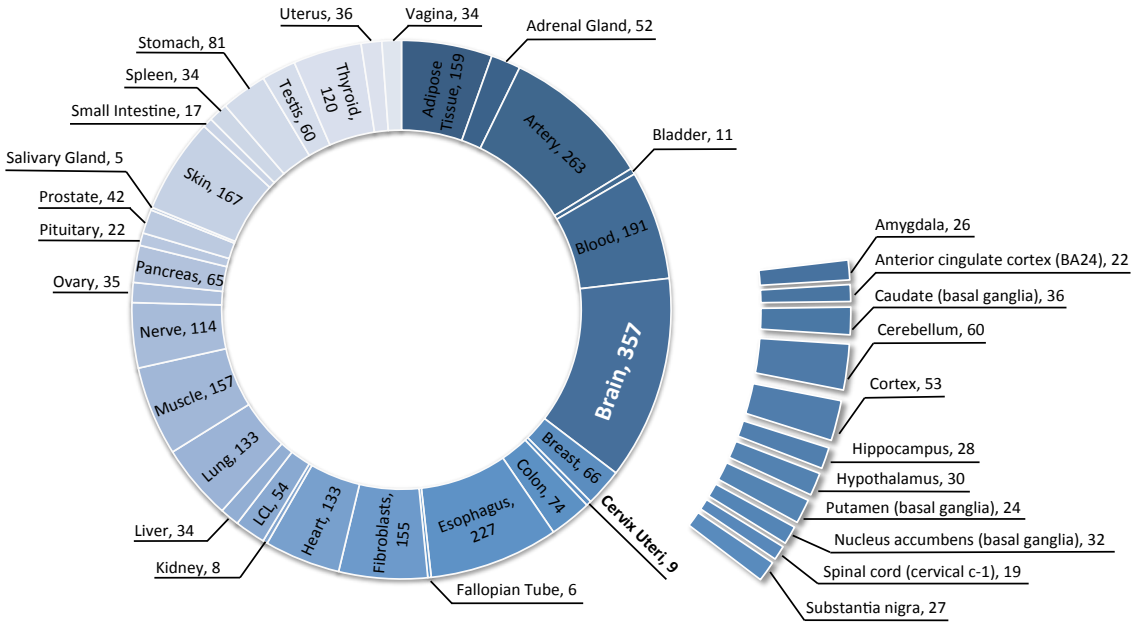


Figure 2.2 GTEx V4 release: sample counts by tissues.

2.2 Consensus MTM detection

2.2.1 Gene and tissue selection

We first evaluated the number of non-detected genes, defined as genes with less than 1 count per million of mapped reads (CPM) in more than 20% of samples in each tissue. We then set, as minimum requirement to construct tissue-specific co-expression networks, ten samples. The number of samples was estimated per tissue, taking into account the maximum combination of outliers (samples in which the number of reads overlapping a gene lie over 1.5 Interquartile Range (IQR)

below the first quartile (Q1) or above the third quartile (Q3) of the tissue raw count estimates of that gene) between couples of detected genes, within the same tissue.

To keep the largest number of samples, we explored gene expression similarity among tissues, in order to find, for those tissues that did not reach the sample threshold, the closest neighbors in terms of body location, function, and gene expression pattern. The genealogy of tissues was estimated, evaluating both whole-genome and traffickome expression patterns. Raw counts were first normalized to render sample libraries comparable despite the differences in depth. The TMM normalization method, implemented in the edgeR R-package, was used to compute a scaling factor for each library. Normalized counts were converted in RPKM units (RPKM= reads per kilobase per million mapped reads), calling the rpkm function within the same package. The function requires in input the vector of gene lengths in bases, which was obtained from the GENECODE (Harrow et al., 2012) release 18 (https://www.gencodegenes.org/releases/18.html). We then calculated, for each tissue, the centroids of gene expressions by obtaining, for each gene, the median expression across all the samples of a given tissue. Finally, hierarchical clustering was performed using different settings: RPKM values were used in log2-transformed ($log_2(1 + RPKM)$) scale; distances between tissues were defined as:

$$D_{ij} = (1 - corr(t_i, t_j))/2, \qquad (2.1)$$

where t_i and t_j are the vectors of gene expression centroids of tissue i and j, and *corr* is the Pearson correlation between them; average-linkage method was applied as clustering strategy. Lastly, genes detected only in the removed tissues were excluded from further analysis.

2.2.2 Network construction

Networks were constructed using WGCNA. For each tissue, we computed an adjacency matrix as:

$$a_{ij} = |corr(g_i, g_j)|^{\beta}, \quad (2.2)$$

where g_i *and* g_j are the expression vectors of gene i and j, and $\beta = 10$ is the soft-thresholding power. Outliers were removed, as described in section 2.2.1, before computing the correlation coefficients.

To reduce features to membrane trafficking genes, only, we assessed the similarity of the correlation distributions of both the coding genome and traffickome by using the Wilcoxon rank-sum test (Wilcoxon, 1946). The feature-reduced adjacency matrices were then transformed into Topological Overlap Matrices (TOMs; see section 1.4.2.3).

2.2.3 Tissue MTM detection and comparison

For each tissue, the corresponding TOM was converted to a dissimilarity measure (1-TOM) and clustered using average-linkage hierarchical clustering. Modules were derived from the dendrogram using the cutreeDynamic function, implemented in the dynamicTreeCut R-package (Langfelder et al., 2008b); function parameters were set to default values, except for the minimum cluster size increased to 30. In order to compare tissue modules, we calculated the gene overlap for each pair of MTMs and used the Fisher's exact test (Fisher, 1922) to assign them a p-value. False Discovery Rates (FDRs) were computed from p-values using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), and the 1% significance level was applied to convert the Fisher matrix into a binary format. Finally, we computed the Jaccard distances among the modules and we

clustered them using average-linkage hierarchical clustering in combination with the cutreeDynamic function (deepSlit=4).

To establish preservation of connection patterns between the modules, we computed, for each gene, the Pearson correlation of its expression with the module eigengenes (1st principal component of the expression matrix of the corresponding module), known as module membership value (kME). MTMs were then clustered by similarities of their network topologies, performing a k-mean clustering (MacQueen, 1967) on the matrix of module kME correlations. Since we wanted to compare this clustering solution to the one obtained matching the module genes, we set the number of groups (k), in which to partition the modules, equal to 5.

2.2.4 Consensus analysis of tissue expression data

To identify consensus MTMs, we considered only trafficking genes detected in at least 22 out of 25 tissues. TOMs were scaled such that their 95th percentiles equaled the 95th percentile of the fibroblast TOM, taken as reference. A consensus TOM was then calculated by taking the component-wise (parallel) 50th percentile of the scaled TOMs, which were first modified assigning zero to the missing connections. Clustering was obtained in an analogous way to tissue TOMs, and modules were derived from the cutreeDynamic function, using the Hybrid method, with a minimum cluster size of 5 and a medium sensitivity (deepSplit=2) to cluster splitting. The "deepSplit" and the "method" parameters were established by evaluating, for each pair of possible values, the mean cluster size, the number of unassigned genes and the composite preservation statistic ($Z_{summary}$) (Langfelder et al., 2011), defined as:

$$Z_{summary} = \frac{Z_{density} + Z_{connectivity}}{2}. \quad (2.3)$$

Each Z-statistic was the result of a pairwise comparison between a given reference (the consensus TOM) and a test set (the scaled TOMs); mean and standard deviation were obtained shuffling, a hundred times, the gene labels in the test network and computing, for each randomized solution, the corresponding permuted statistic. Permutation Z-scores were then calculated as follow:

$$Z = \frac{observed - mean_{permuted}}{sd_{permuted}}. \quad (2.4)$$

For each reference-test pair, and for each module q, defined by the couple of parameters in evaluation, we calculated the $Z_{density}$ as the Z-score of q mean adjacency ($Z_{meanAdj}$) in the test network, and the $Z_{connectivity}$ as the mean Z-score of both q adjacency and intramodular connectivity correlations between reference and test network ($Z_{cor.Adj}$ and $Z_{cor.kIM}$), with:

$$kIM_i^{(q)} = \sum_{\substack{j \in q \\ j \neq i}} a_{ij}^{(q)}. \quad (2.5)$$

The overall median of the $Z_{summary}$ scores, across different modules and test networks, was then used as an indicator of module preservation.

At last, we related consensus modules to tissue modules calculating their gene overlap and using the Fisher's exact test to assign them a p-value; p-values were then adjusted for multiple testing, applying the Benjamini-Hochberg correction.

2.3 Consensus MTM validation

2.3.1 Internal validation

2.3.1.1 Internal validation using module preservation statistics

To estimate internal validity of consensus MTMs, we considered both network-dependent and independent preservation statistics. Network-based statistics ($Z_{meanAdj}$, $Z_{cor.Adj}$, and $Z_{cor.kIM}$) were computed pairwise between the reference and the test set for each consensus module (see previous section); to associate them a

measure of significance, p-values were determined from individual Z-scores under assumption of normality and corrected for multiple testing, applying the Bonferroni FDR-controlling procedure (Bland et Altman, 1995). Finally, to assess the validity of the consensus MTMs, we verified that the median of the top 50% module-wise FDR, across the test set, was below the significance level of 0.05.

In the same way, we computed the Z-scores and the corresponding FDR for the co-clustering network-independent statistic (Langfelder et al., 2011). The co-clustering of module q between the reference and the test set was calculated as the proportion of gene pairs of module q in $Cl^{[ref]}$ that clustered together in $Cl^{[test]}$, denoting with $Cl^{[ref]}$ and $Cl^{[test]}$, the module assignments of the reference and the test set:

$$coClustering(q, q') = \binom{n^{qq'}}{2} / \binom{n^{q}}{2}, \tag{2.6}$$

where

$$\binom{n^{q}}{2} = \frac{n^{q}(n^{q}-1)}{2}, \tag{2.7}$$

is the number of gene pairs of module q in $Cl^{[ref]}$, n^{q} is the number of genes belonging to module q, and $\binom{n^{qq'}}{2}$ is the number of gene pairs of module q that are part of the same module in $Cl^{[test]}$. Validity was established identically to network-dependent statistics.

2.3.1.2 *Graphical representation of two-dimensional distribution of tissue samples*

Principal component analysis was executed on the expression matrices of either the internally validated module genes or the whole traffickome to visualize the distribution of the tissue samples in a two-dimensional space. The Euclidean distance between each pair of samples in the plane was then used to compute the silhouette widths (Rousseeuw, 1987) of the blood and brain samples.

For each sample i, the silhouette width $s(i)$ is defined as:

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))}, \quad (2.8)$$

where $a(i)$ is the average distance between i and all the other points of the cluster to which i belongs, and $b(i)$ is the lowest average distance of i to any other cluster, of which i is not a member.

2.3.1.3 Knowledge-based assessment of module gene functional associations

The co-expression links among the MTM genes were evaluated for functional association by performing a two-sample Wilcoxon rank-sum test between the adjacency distribution of the MTMs that passed the internal validation, and a null distribution, obtained from the adjacency values of the remaining trafficking gene couples; the adjacency distributions of three membrane trafficking complexes (COPI, COPII and ESCRT-III complexes), were used as positive controls and compared with the null distribution as well.

2.3.2 External validation

To consolidate functional inferences, we examined the probability of the MTM genes, to functionally interact, in GIANT and STRING database.

2.3.2.1 The GIANT networks

The Genome-Scale Integrated Analysis of Networks in Tissues (GIANT) (http://giant.princeton.edu) is a webserver that provides an interface to genome-wide functional interaction networks for 144 human tissues and cell types (Greene et al., 2015). These networks are developed using a data-driven Bayesian methodology that integrates thousands of diverse experiments, spanning tissue and disease states**,** contained in more than 14.000 distinct publications. Gene connectivity scores are obtained, automatically weighting each data set for its relevance to each of the tissue- and cell lineage–specific functional contexts, and

calculating the tissue-specific posterior probability of a functional relationship between each pair of genes. The network of functional interactions across all the tissues and cell lines is also provided.

2.3.2.2 The STRING database

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; von Mering et al., 2003) database (http://string-db.org) stores protein-protein interaction data on multiple species, including direct (physical) as well as indirect (functional) associations. These interactions are derived from different "channel", i.e. high-throughput experimental data, co-expression, mining of databases and literature, and predictions based on genomic context analysis. For each channel, protein-protein association scores, scaled between zero and one, are provided. They indicate the estimated likelihood that a given interaction is biologically meaningful, specific and reproducible, given the supporting evidence. Combined scores are also computed by combining the probabilities from the different evidence channels, correcting for the probability of randomly observing an interaction.

2.3.2.3 Validation of functional associations using GIANT and STRING data

The GIANT "all-tissue" network and the version 10.0 of the human STRING database (Szklarczyk et al., 2015) were selected as data sources. Validation of gene functional associations was achieved comparing the GIANT posterior probabilities or the STRING combined scores of the MTM genes with those of the residual trafficking gene pairs. The two-sample Wilcoxon rank-sum test was run to assign the comparisons statistical significance. The distributions of the internally validated MTMs were obtained considering the signed kME of their genes. To determine whether the gene kMEs were statistically significant, a correlation test,

under the null hypothesis to find the current value when the correlation coefficient is in fact zero, was performed. Only module genes with significant kME (fdr<0.05) were retained and only interaction probabilities between genes with the same kME sign (positive or negative) were considered.

2.4 Consensus MTM transcriptional regulation

2.4.1 Regulatory network construction

To enable the inference of putative transcription factors (TFs) involved in the regulation of the module genes, we constructed a target gene-to-TF regulatory network, starting from the data stored in the Transcription Factor Target Gene Database (Plaisier et al., 2016).

This database was constructed by searching, in the promoter regions of 18.153 genes (±5Kbp from the transcriptional start site (TSS)), DNA sequences that significantly matched (FIMO p-value = 1×10^{-5}) one of the 2.331 unique DNA recognition motifs for 690 TFs, collected from JASPAR (Mathelier et al., 2014), TRANSFAC (Matys et al., 2006), UniPROBE (Newburger and Bulyk, 2009) and SELEX (Jolma et al., 2013). DNase I hypersensitivity footprints across 41 diverse cell and tissue types (Neph et al., 2012), were used to provide information about which regions of a promoter had open chromatin, and only motif instances overlapping DNase I footprints were retained.

Of the predicted motif instances, we selected those found in the proximal promoter (±500bp from the TSS) of genes, detected in at least 22 out of 25 GTEx tissues. Instances of TFs, not found among the detected genes, were removed as well. These genes and TFs represented the node pool of a direct graph, in which edges, having a gene/TF as a start-point and a TF as an end-point, indicated that at least one instance of the TF was found in the proximal promoter of that gene/TF. We

also assigned a weight to each edge of the graph, by normalizing the number of instances of a specific TF, found in the promoter region of the gene/TF pointing to it, for the total number of instances of that TF.

2.4.2 Identification of putative transcriptional regulators

Putative TF were identified employing a Random Walk with Restart on the regulatory graph. The function, implemented in the RANKS R-package, computes, for each TF node, the probability of being visited starting from a node/node set, given in input. It also enable to specify a restart parameter (γ), which expresses the probability of restarting from the node/node set after each iteration of the algorithm; we set γ=0.1.

To associate a probability of self-regulation to the starting nodes, in case of TFs, the regulatory network was duplicated with the primary edges pointing to the new network nodes (Fig. 2.3). In this way, the ancestral nodes were only used as starting points of the walk.

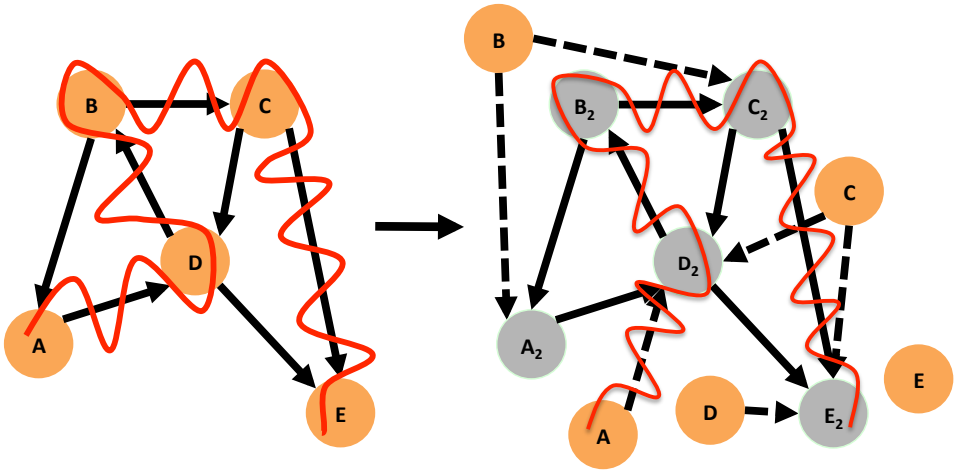


Figure 2.3 Representation of a random walk on a network before and after its duplication. After duplication, the primary edges (dashed lines) point only to the new nodes (gray spheres).

A probability vector, standardized on the maximum value and representing the impact of each TF in the network over the regulation of the starting nodes, was then calculated for each module, starting the random walk from the module genes with positive kMEs; a second run of the algorithm was performed using all the other consensus genes as starting set. The ratio between the two vectors (background correction) defined the putative TFs involved in the specific regulation of that particular module.

2.4.3 Evaluation of TF predictions

We systematically evaluated the sensitivity of the method by comparing TF predictions, made by starting the algorithm from the target genes of 180 ChIP-seq experiments across 5 cell lines and covering 54 different TFs, with those obtained by normalizing the previous for the predictions of the complementary sets; only genes targeted on their proximal promoters were considered for the evaluation.

Specificity was determined by correlating the probability vectors associated to each ChIP-seq experiment.

2.5 Consensus MTM functional annotation

2.5.1 Functional annotation of MTM8

A putative biological role of the consensus MTM8 was determined by analyzing the processed RNA-Seq data of engineered human inducible fibroblasts (hiF-T) together with hiF-T-derived reprogramming intermediates and hiF-T-derived induced pluripotent stem cells (IPSCs), downloaded from Cacchiarelli et al., 2015. Trafficking genes with a dynamic expression were identified applying a Gaussian process (GP) regression method (Kalaitzis and Lawrence, 2011), implemented in the gprege R-package, on the FPKM (FPKM= fragments per kilobase of transcript per million mapped reads) expression matrix, and selecting genes with a log-ratio

of marginal likelihoods >3. WGCNA (β=10; minModuleSize=5; method="hybrid"; deepSplit=2) was employed on the FPKM matrix of the differentially expressed trafficking genes to identify the major dynamic expression patterns during the human cellular reprogramming. We applied two complementary approaches to interpret these clusters: (1) a gene set enrichment analysis for condition-specific gene signatures obtained from human embryonic stem cells (hESCs) in their undifferentiated state or upon differentiation toward the three major germ layers; and (2) a comparison with cluster of differentially expressed genes during reprogramming, annotated for developmental cell identity (Edgar et al., 2013).

In the first approach, we produced an ordered list of genes based on their specificity for a given condition (hESCs in pluripotent conditions or differentiated into progenitor populations of the early embryonic germ layers, whose RNA-Seq data were published in Gifford et al., 2013) by using the *csSpecificity* function in the cummerbund R-package. The function computes for each gene, in each condition, a specificity score *s*, defined as:

$$s_{g,i} = 1 - JSD(p_g, q_i), \tag{2.9}$$

where JSD is the Jensen-Shannon distance, p_g is the expression profile of a given gene g expressed as a density (probability) of $\log_{10}$(FPKM+1), and q_i is the unit vector of 'perfect expression' in a particular condition i. We then performed a GSEA to evaluate whether the members of the clusters tended to occur toward the top of the list, in which case they were correlated with that particular condition.

In the second approach, we calculated the gene overlap between the clusters of trafficking genes differentially expressed during reprogramming and those identified in Cacchiarelli et al., 2015. To assign the comparison statistical

significance we computed the Fisher's exact test and corrected the p-values adopting the Benjamini-Hochberg correction.

In the same way we compared the reprogramming clusters to the consensus MTMs.

2.5.2 Functional annotation of the other consensus modules

To functionally annotate the other consensus modules, we adopted a two-step approach. In the first step, from the GIANT all-tissue network, we selected significant interactions involving the module genes; in particular, only the interactions involving module genes with positive kME were considered, including those between these genes and non-trafficking genes. An interaction was deemed significant if the corresponding posterior probability was among the top 1% of the total posterior probabilities. In order to assess cellular processes enriched within the consensus MTMs, we collected gene sets from both the C2 and C5 collections of the Molecular Signatures Database (MSigDB, Subramanian et al., 2005). The hypergeometric test was then applied using as 'balls drawn from the urn' the module interactors, as 'white balls' the genes of the collections, and as 'black balls' all the other genes; processes significantly enriched (fdr < 0.05) were used to annotate the modules.

In the second step, Drug-Set Enrichment Analysis (DSEA, Napolitano et al., 2016) was carried out to confirm the predictions of the first step. The method is able to identify cellular processes that are targeted by a set of drugs, which, in accordance with the Mode of Action by NeTwoRk Analysis (MANTRA) tool (Iorio et al., 2010), were chosen for their ability to up- or down-regulate the module genes. This step was applied only when among the top 20 drugs (ranked by absolute enrichment scores), we could identify a subset of at least 4 drugs that were able to up or down regulate exclusively the genes of a specific module. In such a case, we compared the

processes, which came out significant (fdr < 0.05) in this step, with those that had been found to be significant in the first step.

Consensus module enrichments for C2 (KEGG gene sets) and C5 (GO Molecular Function sets) annotations are shown in supplementary Figures S1-S23.

2.6 Tissue-specific MTM detection and validation

2.6.1 Identification of tissue-specific MTMs

To identify tissue-specific MTMs, we combined differential correlation with weighted network analysis. In this regard, we considered all the correlation between trafficking genes detected in at least one tissue, assigning zero to missing coefficients. In order to establish the differences in gene correlations between tissues of different sample sizes, Pearson correlation coefficients (r) were converted into Z-scores using Fisher's Z-transformation:

$$Z = \frac{1}{2}\ln\left(\frac{1+\mathrm{r}}{1-\mathrm{r}}\right). \tag{2.10}$$

The matrix of Z-differences was then defined as:

$$\Delta Z_{ij} = \frac{Z_i - Z_j}{\sqrt{(N_i-3)^{-1}+(N_j-3)^{-1}}}, \tag{2.11}$$

where i denotes the reference tissue, j all the tissues other than reference, and N_i and N_j their respective sample sizes.

Tissue-specific MTMs were identified from the set of matrices resulting from the Z-differences between a reference and each of the other tissues, in four steps. First, ΔZs were converted into adjacency matrices, calculating the two-sided probabilities corresponding to the Z-difference scores, and then raising the complements of these probabilities to the power of ten; second, the adjacency matrices were transformed to TOMs, which were normalized by equalizing their 95th percentile; third, a consensus TOM was calculated by taking the parallel 50th

percentile of the scaled TOMs and was converted into a dissimilarity measure (1-TOM); forth, average-linkage hierarchical clustering was applied followed by cutreeDynamic cutting. In this case, cutreeDynamic parameters (method="tree", minModuleSize=5 and deepSplit=TRUE) were evaluated for the mean cluster size and the number of the emerging tissue-specific MTMs.

Finally, to assess the accuracy of the method, we checked the differential co-expression of the synaptic vesicle genes by looking at their adjacency scores in the consensus TOM of each tissue; a Wilcoxon rank-sum test was then performed to determine if the brain distribution was statistically higher than the combined distribution from all other tissues.

2.6.2 Internal and external validation of tissue-specific MTMs

For the internal validation of tissue-specific MTMs, we used almost the same criteria applied to the consensus modules. Network-dependent ($Z_{meanAdj}$, $Z_{cor.Adj}$, and $Z_{cor.kIM}$) and independent ($Z_{coClustering}$) statistics were computed pairwise between the reference and the test set for each tissue-specific module, and p-values were determined from individual Z-scores under assumption of normality. Unlike consensus MTMs, tissue-specific modules were considered internally validated if the 50% of the module-wise p-values, across the test set, was below the significance level of 0.05.

Along with the preservation statistics, we also evaluated the differential co-expression of the internally validated modules in GIANT. From GIANT, we selected functional interaction networks for the 19 human tissues that were found to be in common with the GTEx tissue categories. The significance of differential co-expression of the tissue-specific MTMs was then assessed by performing a two-sample Wilcoxon rank-sum test between the GIANT posterior probability

distributions of the module gene pairs in their respective network and the combined distributions of those gene pairs from all other networks. Before the comparison, the distributions were converted to the corresponding standard normal forms.

2.7 Normalization of random walk scores to exploit network information

2.7.1 Co-occurrence network of biomedical concepts

The analysis of the network of biomedical concepts provided by Grammatica et al., 2014, was part of a side project, started during the first year of my PhD. The network hosts 24,984 entities (nodes of the network, e.g., drugs, diseases, genes), organized into 12 classes (Table 2.1 reports entity cardinality per class), which are linked on the basis of literature mining. In order to score associations between entities not directly connected in the network, a random walk (RW) is computed between each pair of nodes in the network. The random walk value between node A and node B is thus directly proportional to the amount of literature forming the association path starting from node A and ending at node B.

Table 2.1 Cardinality per class of entity.

CLASS	Number of Elements
CELL	73
DISEASE	7,474
DRUG	2,938
ENZYME	158
GENE	13,060
HORMONE	65
ORGAN	21
PEPTIDE	132
PROCESS	15
PROTEIN	809
RECEPTOR	219
OTHER(*)	20
TOTAL	**24,984**

(*) Miscellaneous elements used only for disambiguation purposes

2.7.2 Normalization procedure

As there were nodes in the network with the tendency to have high RW scores and, usually, these were nodes associated with a considerable number of scientific manuscripts, we reduced the number of times these nodes were selected by a simple cut-off procedure, applying the following formula to normalize RW scores:

$$normRW_{ij} = \frac{l_j * RW_{ij}}{sum_j * N}, \quad (2.12)$$

where l_j are the number of non-null RW scores reaching node *j* (starting from any node of a given class, i.e., drugs, diseases, etc.), $RW_{ij}$is the RW value associated to the pair ij, sum_j is the sum of all RW values reaching node *j* and starting from any of the nodes of a given class, and *N* is the number of nodes associated to a given class. The procedure described was used to normalize drug-drug and disease-drug RWs, i.e., RWs that had a drug as a source and destination, and a disease as source and a drug as destination, respectively.

2.7.3 Evaluation of network predictive performances after score normalization

2.7.3.1 Goodness of disease-drug and drug-drug associations

Both the un-normalized and normalized versions of the networks were compared against golden standards to evaluate their goodness. Specifically, the disease-drug network was first transformed in a disease network by computing the Pearson correlation coefficients between the drug-profiles associated with each disease and then compared to the Online Mendelian Inheritance in Man (OMIM)-derived network, whereas the drug-drug network was directly compared with the anatomical therapeutic chemical (ATC)-derived drug network (see section 2.7.3.2 and 2.7.3.3 for details about the construction of the golden standards). Similarity scores within the networks were ordered by significance and employed to derive

true positive rate (TPR) and false positive rate (FPR) at various threshold settings. ROC curves were created by plotting the true positive rate (TPR) against the false positive rate (FPR), and the corresponding AUCs were calculated. Finally, AUC significances were conveniently computed via the Mann-Whitney U statistic (Mann and Whitney, 1947; Tang and Balakrishnan, 2011).

2.7.3.2 ATC-derived drug network

The ATC classification system (https://www.whocc.no/atc/) is adopted for the classification of drugs based on different characteristics such as, their therapeutic and pharmacological properties, and the way they act on target tissues/organs. Each drug entry in the classification system has five ATC codes associated, one for each level of the ATC hierarchy. For our comparisons we considered the 4th level of the classification that covers anatomical, therapeutic, pharmacological and chemical properties. The ATC drug network was built by connecting drugs sharing similar ATC level-4 codes. Hence, two drugs with similar anatomical, therapeutic, pharmacological and chemical impact were linked by an edge in the ATC drug network. The final ATC drug network contained 559 nodes, in common with the co-occurrence network, connected by 1014 edges.

2.7.3.3 OMIM-derived disease network

OMIM (Amberger et al., 2009) is a compendium of human genes and genetic Mendelian phenotypes. Technically, OMIM can be seen as a list of pairs where one element is a Mendelian disease and the other a mutated gene, with diseases eventually associated to more than one mutation on the same or different gene. OMIM network was built under the realistic assumption that a mutation in OMIM affects one or more pathways leading to the diseased phenotype. Each disease was thus associated to all genes of the pathways, known to be linked with the

corresponding mutated gene. Then, diseases were related to each other by computing the Jaccard distance. The p-value cutoff over Jaccard significances was set to 0.01. The final OMIM disease network contained 416 nodes, in common with the co-occurrence network, connected by 5310 edges.

Chapter 3

Results

3.1 Consensus MTM detection

3.1.1 Rationale for the choice of GTEx

The Genotype-Tissue Expression (GTEx) Project (GTEx Consortium, 2015) is a resource project designed to evaluate gene expression variability in human tissues. Samples are obtained from harvesting autopsies of relatively healthy individuals, which must pass precise selection criteria:

- 21 ≤ Age (years) ≤ 70
- 18.5 < Body Mass Index (BMI) < 35
- Time between death and tissue collection below 24 hours
- No whole blood transfusion within 48 hours prior to death
- No history of metastatic cancer
- No chemotherapy or radiation therapy within the 2 years prior to death
- Absence of diseases or disorders that would disqualify someone to donate organs or tissues

Thus, GTEx provides a wealth of gene expression data to study the underlying genetics of normal physiology in human tissue, aspects that best fit with the aims of this project.

3.1.2 Removal of certain tissues and formation of new categories

The initial GTEx V4 expression matrix consisted of 2,921 samples, divided in 32 tissues (brain sub-regions were considered as unique tissue), and 55,993 transcribed genes, including protein-coding genes, pseudogenes, and long noncoding RNAs (lncRNAs). As a result of gene and tissue selection (see section 2.2.1), 24 samples b(elonging to bladder, kidney and salivary gland tissues) and

32,221 genes (31,846 of which were not detected at all and 375 detected only in the excluded tissues) were removed (Table 3.1). Cervix uteri, fallopian tube and small intestine tissues, which had not passed the sample selection criteria, turned out to have genome and traffickome expression patterns similar to those of tissues that shared analogous body location and function with them (Fig. 3.1).

TABLE 3.1	N. of samples	Max. Out. Comb.	Residual samples
Adipose Tissue	159	63	96
Adrenal Gland	52	20	32
Artery	263	96	167
Bladder	11	8	3
Blood	191	65	126
Brain	357	151	206
Breast	66	25	41
Cervix Uteri	9	8	1
Colon	74	26	48
Esophagus	227	80	147
Fallopian Tube	6	4	2
Fibroblasts	155	56	99
Heart	133	56	77
Kidney	8	6	2
LCL	54	22	32
Liver	34	16	18
Lung	133	46	87
Muscle	157	53	104
Nerve	114	42	72
Ovary	35	17	18
Pancreas	65	23	42
Pituitary	22	11	11
Prostate	42	20	22
Salivary Gland	5	4	1
Skin	167	67	100
Small Intestine	17	12	5
Spleen	34	16	18
Stomach	81	33	48
Testis	60	21	39
Thyroid	120	41	79
Uterus	36	15	21
Vagina	34	16	18

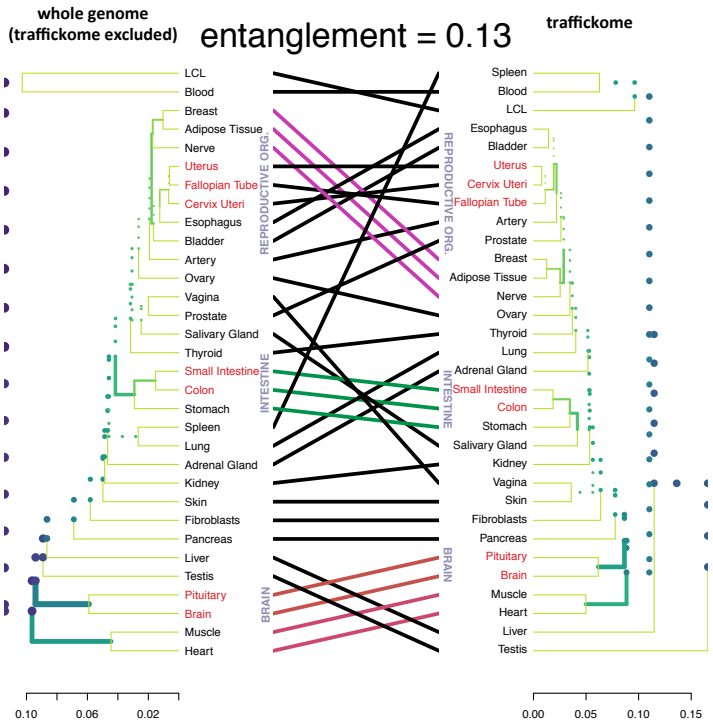


Table 3.1 Tissue selection taking into account the maximum outlier combination between two genes. Some tissues were completely removed (light-red cells) while others (yellow cells) were grouped in more general categories according to a hierarchical clustering procedure (Fig. 3.1). Conditional formatting was used to add data bars to the cells of the second and third column, and to color scale, with different green gradations, the cells of the fourth column. **Figure 3.1 Tanglegram's plot of side-by-side tissue trees**. Trees were obtained using as signature both traffickome and non-traffickome genes. The quality of the alignment of the two trees was measured using the entanglement function in the dendextend R-package; the entanglement coefficient at the top center indicates a good alignment (value close to 0).

Based on this observation, we created new categories: "*reproductive organs*", (comprising cervix uteri, fallopian tube and uterus tissues) and "*intestine*" (formed

by small intestine and colon tissues). In this process we also included the pituitary gland (which was just above the sample threshold limit), found highly similar to brain and for this reason merged with it. The final dataset was hence composed of 23,772 genes and 2,897 samples, rearranged in 25 tissues. Figure 3.2 shows, for each of the selected tissues, how the log2-transformed count distributions appear before and after gene filtering.

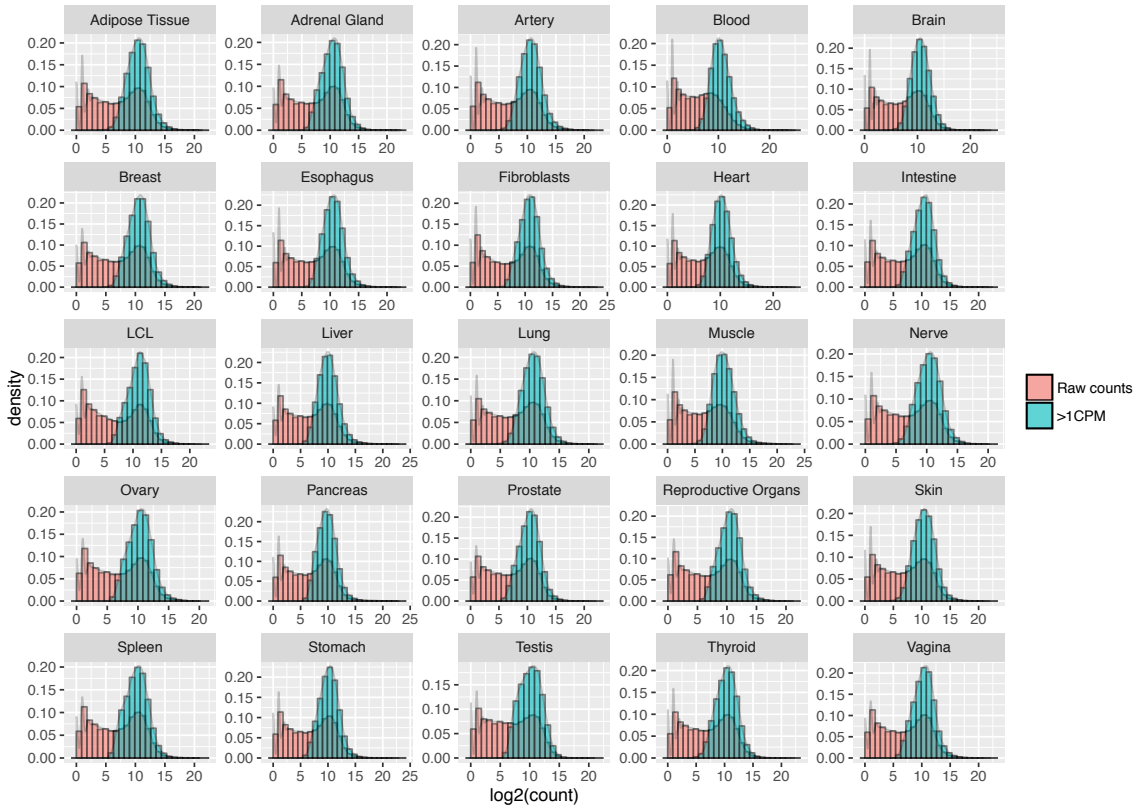


Figure 3.2 Frequency histograms showing log2-transformed count distributions in each tissue. Unfiltered (light red) and filtered (cyan) distributions are shown; after filtering, the distributions assume the aspect of long–tailed Gaussians.

3.1.3 Decision-making process concerning the first two steps of WGCNA

For each of the 25 tissues that passed the selection, we constructed a co-expression network using weighted correlation network analysis (complete

procedure in section 2.2.2). The first two steps of the network analysis provide for the definition of both a similarity measure and an adjacency function.

We estimated the similarity between gene expression profiles by computing Pearson correlation. It is known that Pearson correlation is strongly affected by extreme values, which may exaggerate or dampen the strength of the relationship between two variables, and by the number of paired observations of the related variables; in fact, the smaller the sample size, the greater the likelihood of obtaining a spuriously high correlation coefficient. Thus, before computing the correlation coefficients between pairs of genes, outliers were removed after checking that their removal did not reduce significantly the number of gene expression measurements in tissues with less than 50 samples, i.e liver, ovary, prostate, spleen and vagina. We established that, with an average of 12,733 genes, detected in the tissues listed above, and 63,665 cases (12,733 genes x 5 tissues), 79% of the genes in each of these tissues had at most two outliers, justifying our choice. Figure 3.3 recapitulates the outlier distribution in each tissue.

Correlation coefficients were then converted into adjacency scores, employing a soft-thresholding to power their values. Generally, this threshold is chosen following the scale-free topology criterion (representing the scale-free topology index (y-axis) as a function of the soft-thresholding power (x-axis), is the lowest value at which the scale-free topology index flattens out upon reaching a value above 0.85).

Although most biological networks are close to being scale-free (Xin Hu et al., 2016), there was no motivation for applying the scale-free topology criterion here, as our samples comprised several different tissues. We set, instead, the power equal to ten in order to emphasize stronger associations (larger correlation coefficients) and flatten weaker ones (lower correlation coefficients) (Fig. 3.4).

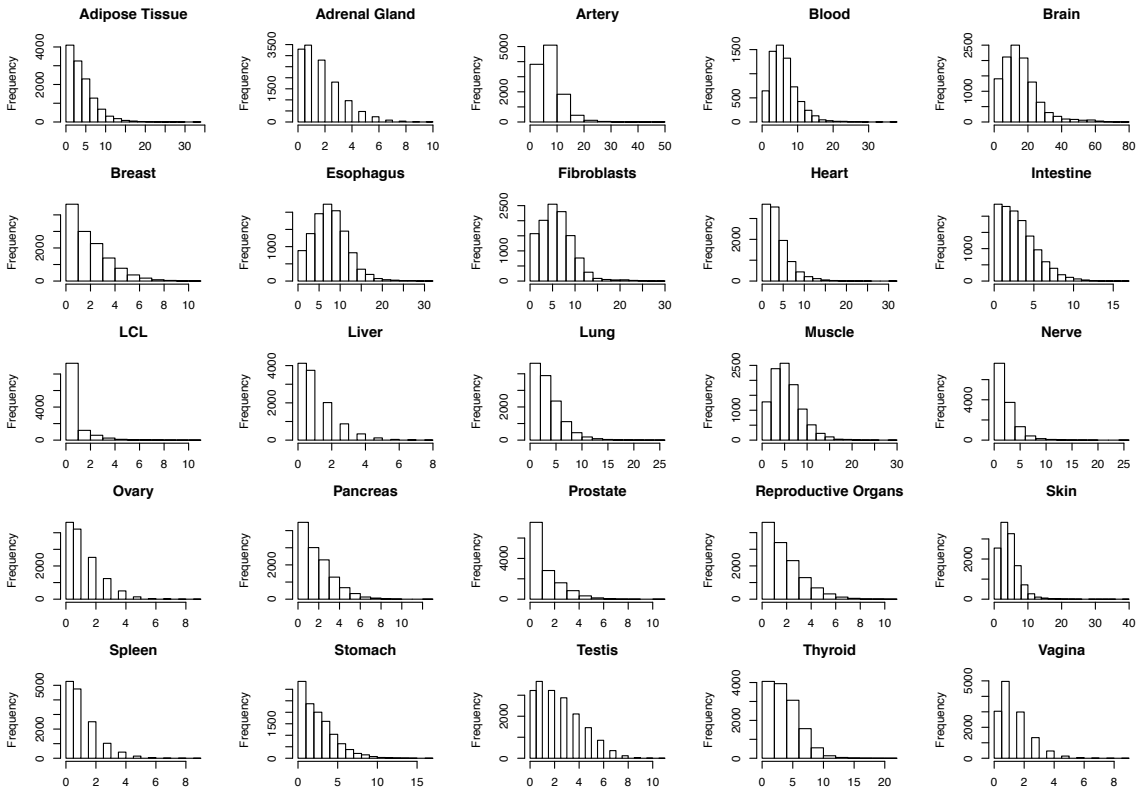


Figure 3.3 Frequency histograms showing tissue outlier distributions.

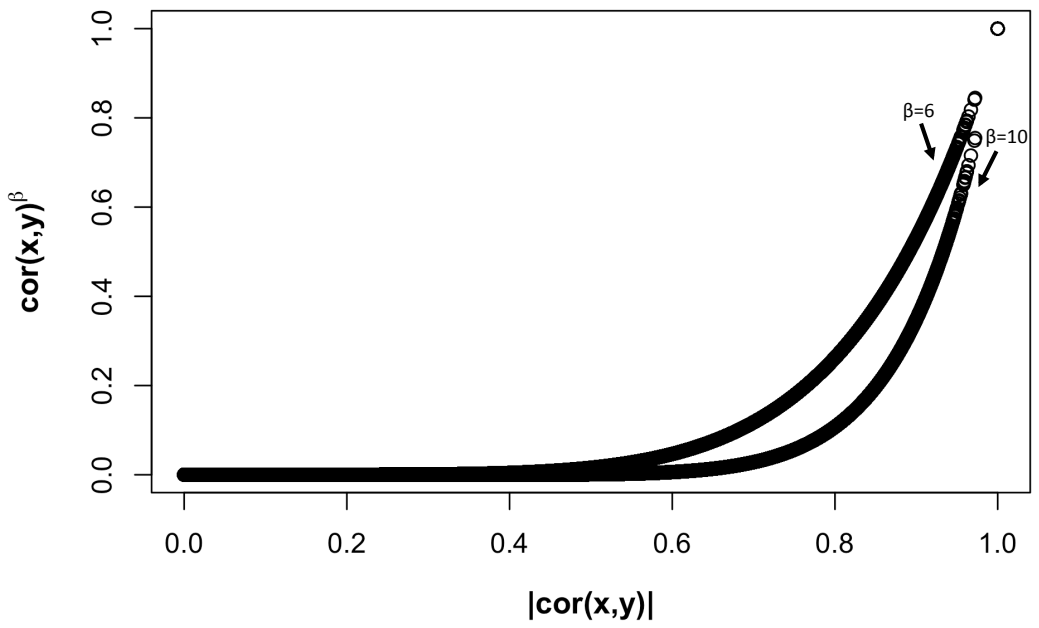


Figure 3.4 Comparison between soft-thresholding powers. It is recommended to choose β=6 so that the resulting co-expression network follows an approximate scale-free topology. Another way is to consider β as a stringency parameter; a high value of β means putting less emphasis on lower correlation coefficients, and therefore being more statistically stringent.

3.1.4 Thesis in favour of considering only trafficking genes

Before proceeding with the next two steps of the network analysis, which are the definition of a measure of node dissimilarity and the identification of the modules, we reduced the set of genes, whose modules were supposed to be identified, to the 1,237 trafficking genes detected in at least one tissue.

This step required some attention; in fact, if the filtering had forced the association between trafficking genes, there would have been a high risk to generate false co-expression modules. Thus, for each tissue we compared the correlation distributions of trafficking and non-trafficking genes using the Wilcoxon rank-sum test; in 24 out of 25 comparisons (the fibroblast dataset was the only exception), we rejected the null hypothesis (p-value<0.05), according to which the traffickome correlation distribution was lower than the other one. This result suggested that the trafficking genes would have clustered together regardless of the removal of the other genes, which, at that point, were eliminated.

3.1.5 Trafficking genes are organized in modules preserved among tissues

One of the purposes of constructing the co-expression networks was to determine if there were groups of trafficking genes that were co-expressed in human tissues. To this end, we identified 121 modules from the selected-tissue networks and matched them based on the number of overlapping genes, as described in section 2.2.3. Table 3.2 shows, for each tissue, the number of trafficking genes detected, the number of modules identified and the average number of genes per module.

We found that 692 out of 7,011 possible matches, involving 109 modules, were statistically significant (fdr<0.01). Cluster analysis revealed the existence of four groups of module similarities; a fifth, labeled 0, consisted of modules not assigned

to any of the previous clusters (no significant matches were observed between these modules and the other ones, Fig 3.5 panel A).

TABLE 3.2	N. of detected genes	N. of modules	Average n. of genes
Adipose Tissue	1095	5	123
Adrenal Gland	1102	5	145
Artery	1080	5	130
Blood	952	5	150
Brain	1124	7	123
Breast	1106	5	153
Esophagus	1089	3	302
Fibroblasts	1042	4	213
Heart	1075	5	163
Intestine	1128	5	187
LCL	999	5	112
Liver	1055	4	196
Lung	1129	4	147
Muscle	1026	6	99
Nerve	1114	4	190
Ovary	1094	6	147
Pancreas	1107	2	276
Prostate	1136	5	184
Reproductive Organs	1106	5	151
Skin	1119	4	186
Spleen	1108	6	130
Stomach	1103	5	175
Testis	1187	7	124
Thyroid	1133	3	233
Vagina	1120	6	152

Table 3.2 Average number of genes per tissue module. Data bars were added to the cells related to the average number of genes.

These preliminary results suggested that the majority of the trafficking genes constituted co-expression communities shared by several tissues, while only a small fraction of these genes contributed to the formation of tissue-specific modules (an entire section on the identification of tissue-specific modules will be presented in paragraph 3.5).

We also evaluated if modules within the same cluster had similar network topologies (see section 2.2.3); again, we clustered the modules and found that 66 out of 109 modules were assigned to the same groups (Fig 3.5 panel B).

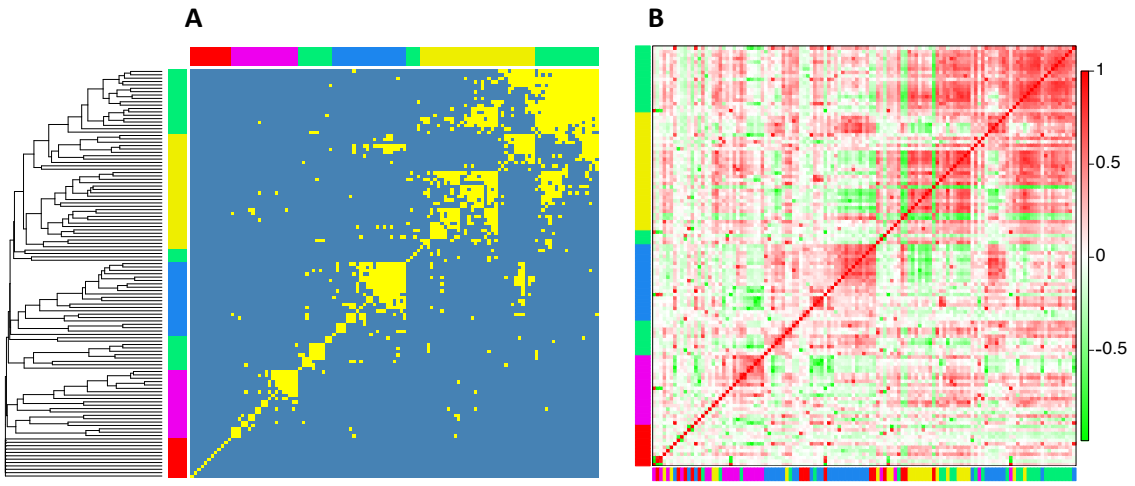


Figure 3.5 Heatmaps depicting the relationships between tissue modules. A In yellow, significant (fdr<0.01) gene overlaps between modules; on the sidelines, the clustering solution obtained performing average-linkage hierarchical clustering in combination with the cutreeDynamic function (deepSlit=4). **B** Module kME correlations; correlation coefficients are colored from green (low coefficients) to red (high coefficients). The result of the k-means clustering is reported along the lower edge.

3.1.6 Definition of core members shared by similar modules

Having established the existence of groups of co-expressed genes common to multiple tissues, we built a consensus TOM, starting from 1,027 trafficking genes detected in at least 22 out of 25 tissues (details in section 2.2.4), in order to determine the exact number and the core members of these groups. Modules were identified by the cutreeDynamic function, whose parameters were explored to find the most acceptable clustering solution (evaluation criteria in section 2.2.4). The parameters evaluated were the *"method"* of adaptive branch pruning of hierarchical clustering dendrograms, and the sensitivity to cluster splitting, *"deepSplit"*. The method is present in two variants: the *"tree"* cut that relies solely on the dendrogram, and the *"hybrid"* cut that improves the detection of outlying members of each cluster. The deepSplit parameter, instead, provides a rough control over sensitivity to cluster splitting. For the *hybrid* method the value is an

integer in the range 0 to 4; for the *tree* method it is a logical value (i.e. TRUE or FALSE). The higher the value (or if TRUE), more and smaller clusters are produced. Using the tree cut method in combination with both of its deepSplit variants resulted in very small modules (8-10 genes on average) with a large number of unassigned genes (>800) (Fig. 3.6 and Table 3.3). However, the modules detected were highly preserved in tissues. The hybrid cut method, conversely, produced modules with different average size and degree of preservation, depending on which deepSplit value was used in combination with it, although the number of unassigned genes was very low (Table 3.3).

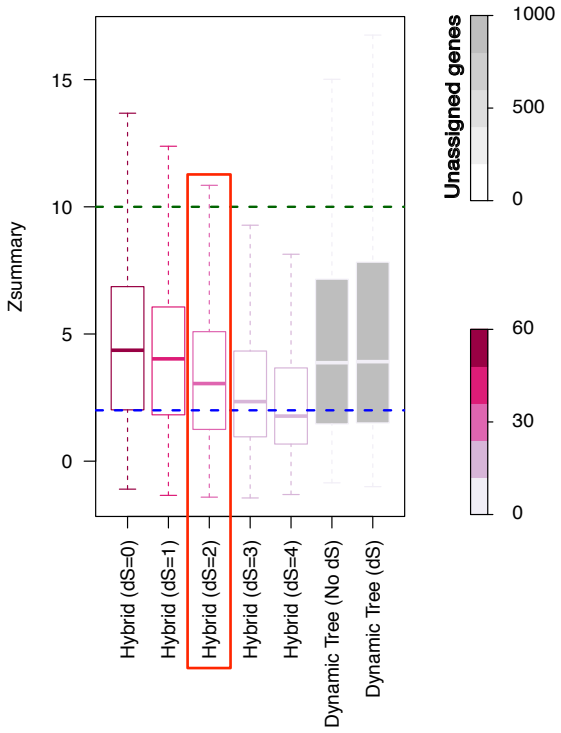


TABLE 3.3	Zsummary	Average module size	Unassigned genes
Hybrid (dS=0)	4.36	59	31
Hybrid (dS=1)	4.02	45	29
Hybrid (dS=2)	3.05	28	25
Hybrid (dS=3)	2.34	19	23
Hybrid (dS=4)	1.77	14	21
Dynamic Tree (No dS)	3.87	10	810
Dynamic Tree (dS)	3.91	8	818

Figure 3.6 Box plots of preservation $Z_{summary}$ scores for modules detected by different tree cut methods. The outlines and bodies of the boxplots are colored with different gradation of violet and gray depending on the average module size and the number of unassigned genes, respectively. The red rectangle indicates the method chosen to identify consensus modules. **Table 3.3 Parameters evaluated for each tree cut method, in numbers.** Values are colored from red (bad scores) to green (good scores).

We decided to use the hybrid method to assign most of the genes and then eventually discard the modules at the validation stage (see section 3.2), and deepSplit equal to 2 to generate modules with a reasonable number of genes to justify a possible biological function (neither too large nor too small); in this way, we identified 36 consensus modules (results in supplementary Table 1), which were related to the modules we discussed in the previous section (Fig. 3.7). Contrary to what was expected, few consensus modules had significant enrichment within a given cluster while others showed a few dispersed matches across all cluster modules.

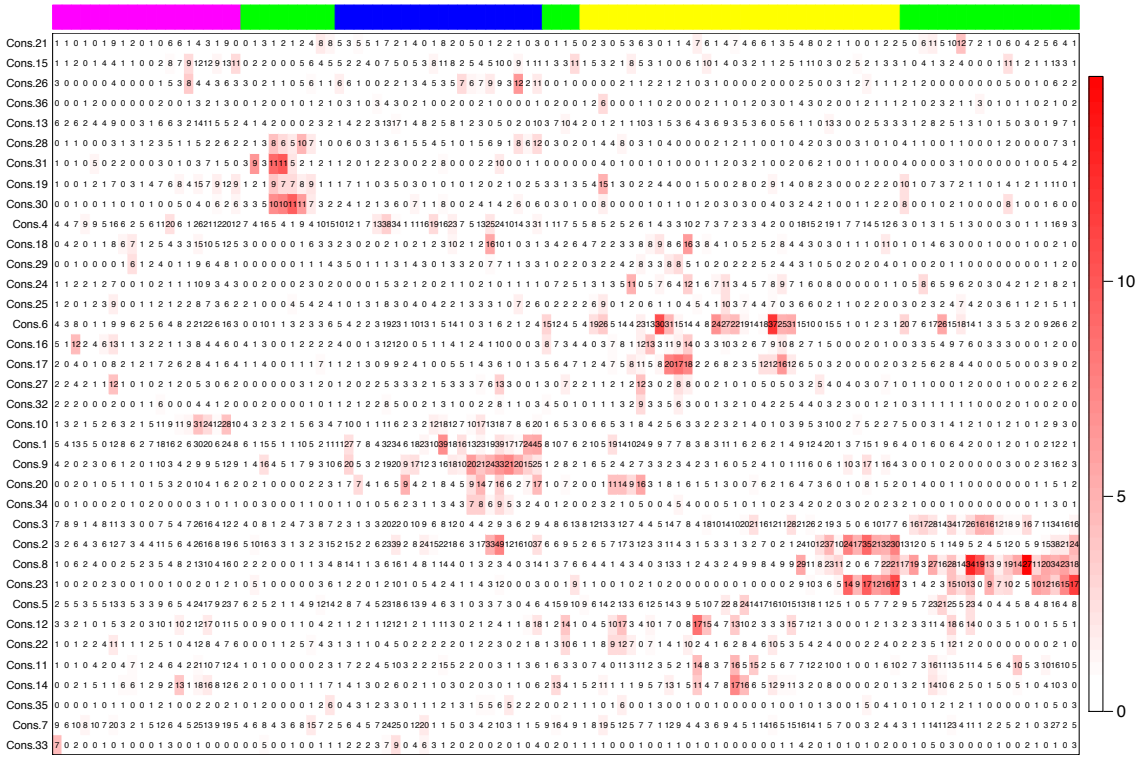


Figure 3.7 Correspondence of consensus modules with tissue modules. Each row of the table corresponds to one consensus module, and each column corresponds to one tissue module (labeled with cluster colors). Numbers in the table indicate gene counts in the intersection of the corresponding modules. Coloring of the table encodes −log(p), with p being the Fisher's exact test p-value for the overlap of the two modules. The stronger the red color, the more significant the overlap.

3.2 Consensus MTM validation

3.2.1 Preservation of consensus modules in individual tissues

The preservation of the consensus MTMs was evaluated in individual tissues; for each double, consensus module-tissue, we calculated three network-dependent, and one network-independent, preservation statistics, as detailed in section 2.3.1.1. We found that 30 modules were significantly preserved (fdr<0.05) in at least seven tissues (the rounded up quarter) for three out of four statistics while only 13 were preserved for all of them (Fig 3.8).

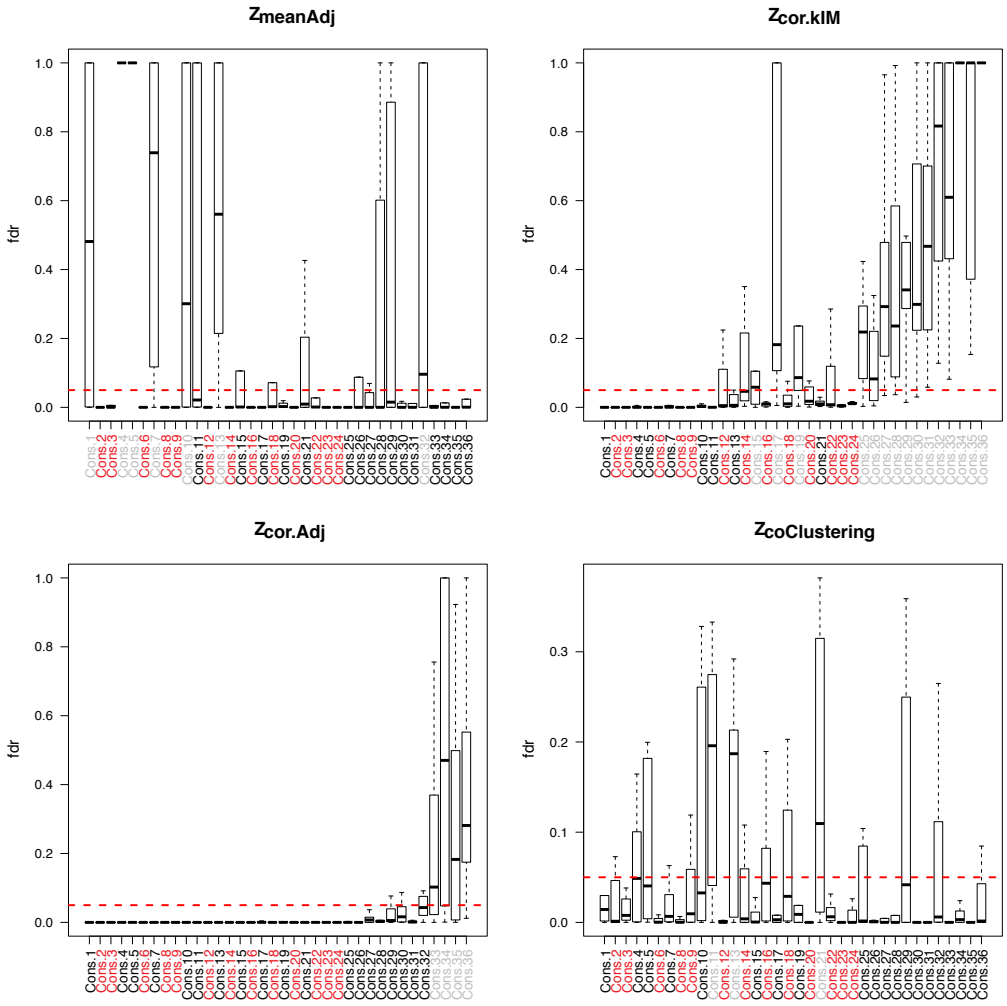


Figure 3.8 Box plots of the top 50% fdr determined from the Z-scores of individual statistics for each reference-test pair. In light red, consensus modules significantly preserved (median fdr<0.05) for all four statistics; in black and light gray, respectively, modules significantly and not significantly preserved for that particular statistic.

From here onwards, the term modules indicates the 13 consensus modules significant for the four preservation statistics.

The heatmaps in Figure 3.9 recapitulate the $Z_{summary}$ and the $Z_{coClustering}$ scores of the modules in each tissue; values > 10 indicate strong evidence that the modules are preserved, values between 2 and 10 indicate weak to moderate evidence of module preservation, while values < 2 indicate no evidence of module preservation (Horvath, 2011). According to this score system, consensus 20 and 23 appeared to be the most preserved modules.

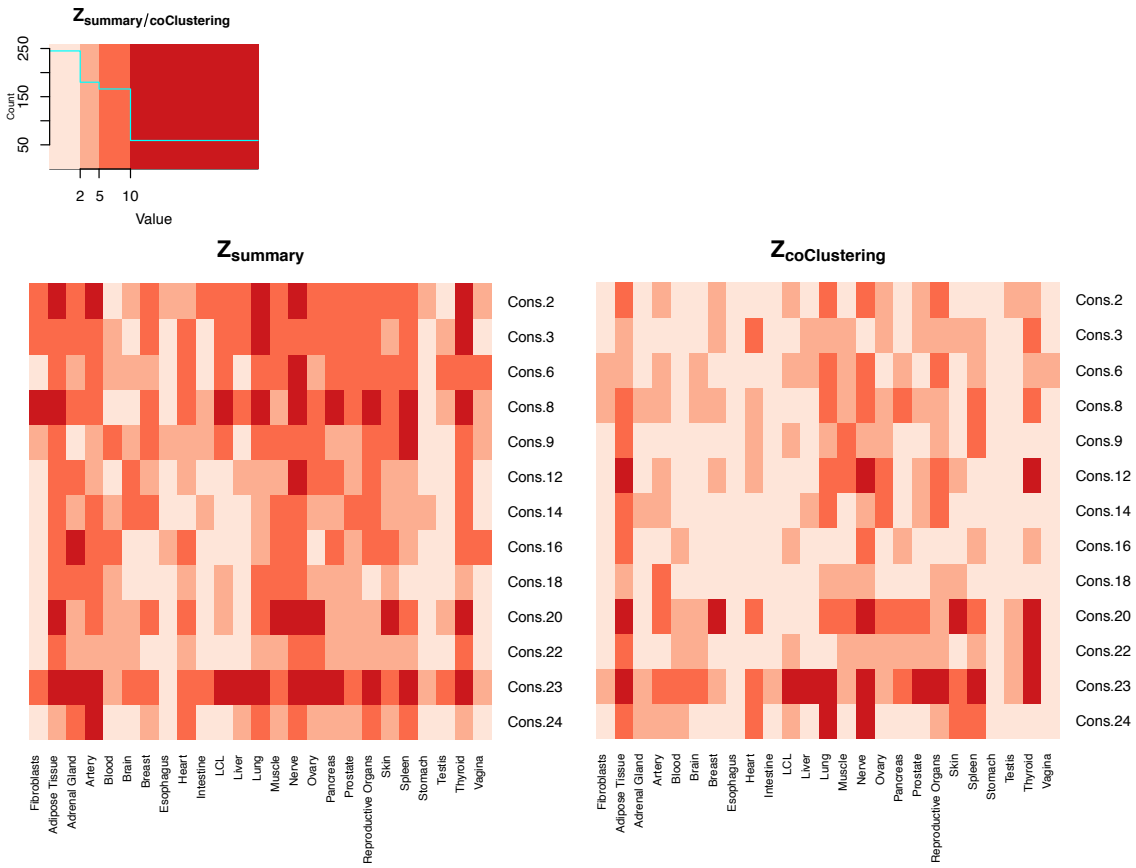


Figure 3.9 Heatmaps showing preservation scores for the consensus module in each tissue. $Z_{summary}$ (left) and $Z_{coClustering}$ (right) scores are colored with different gradation of red; darker reds indicate higher values of preservation.

We then evaluated the distances among the samples by performing a principal component analysis on the gene expression profiles of both traffickome and module genes (Fig. 3.10; section 2.3.1.2 for details).

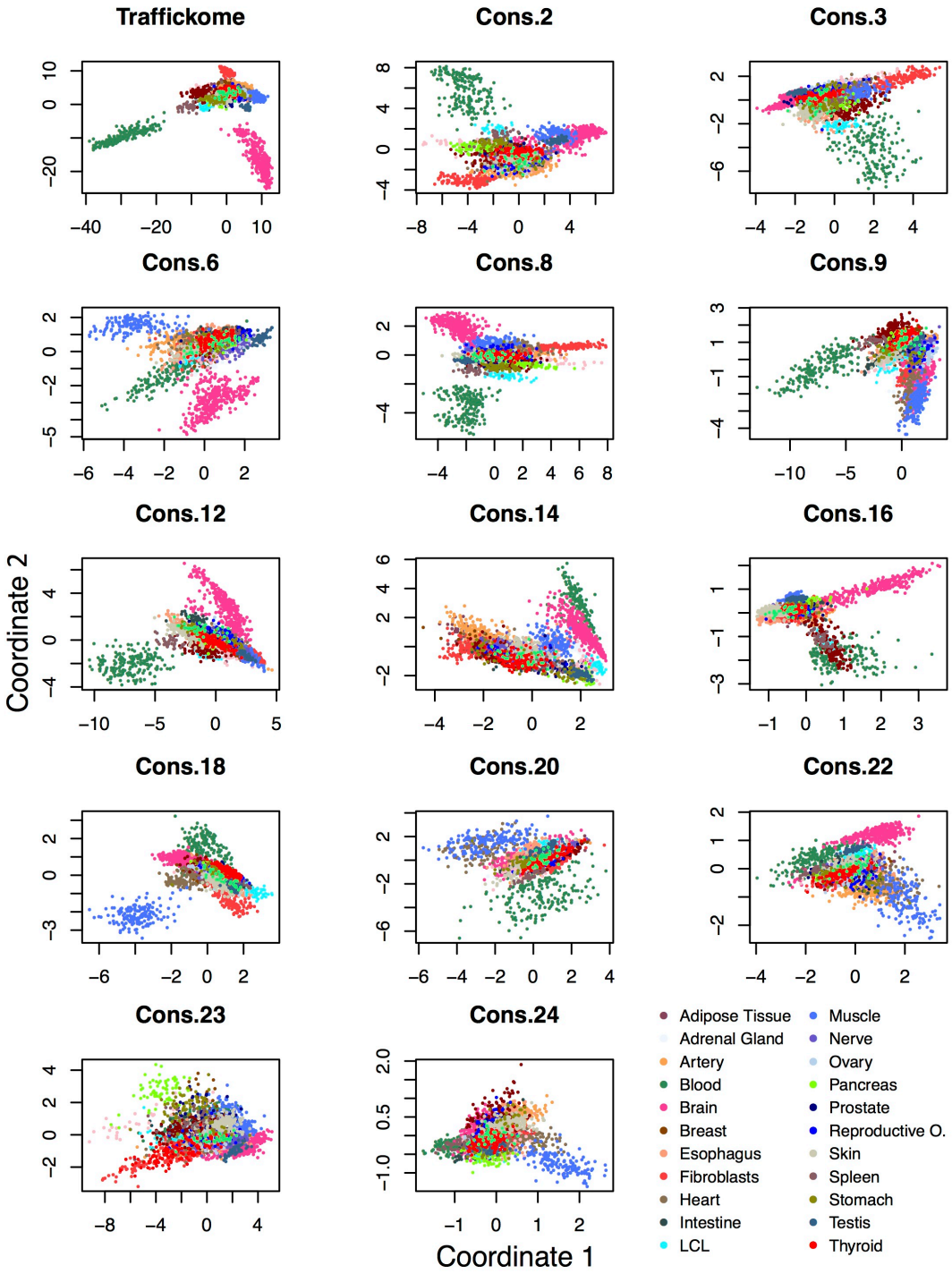


Figure 3.10 Multidimensional scaling of GTEx samples based on traffickome or consensus gene expression. Samples are projected onto the first two principal components.

We expected the sample distances to be lower when using the module genes as signature rather than the entire traffickome, and indeed characteristic traffickome separations between nonsolid (blood) and solid tissues, and within solid tissues, between brain and the other ones, were reduced in preserved modules, as proven by the average silhouette width of their samples (Fig. 3.11); high values (>0.5) indicate that the samples are well matched to their own cluster and poorly matched to neighboring clusters.

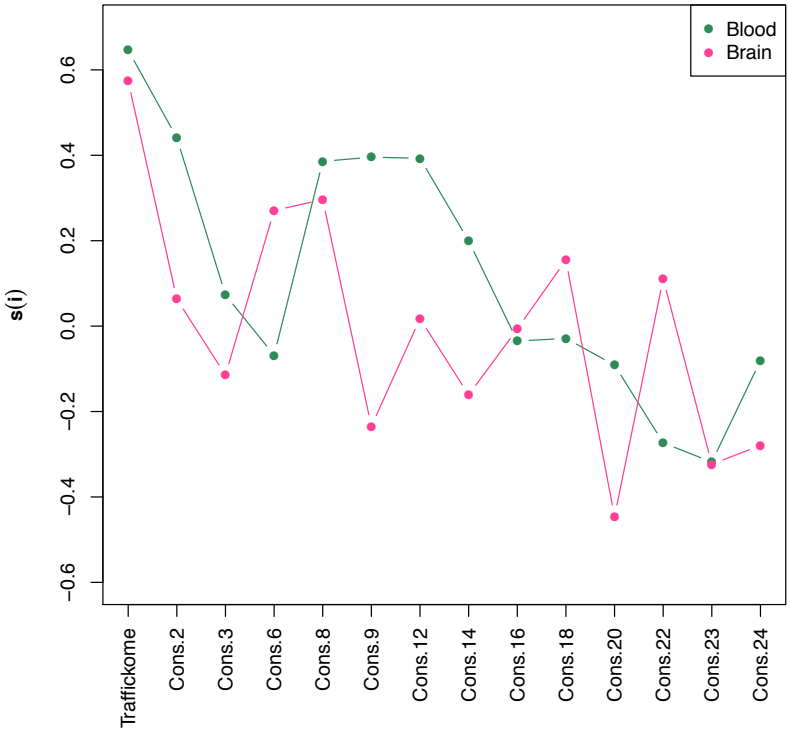


Figure 3.11 Average silhouette widths of brain and blood samples, following PCA. Silhouette coefficients are computed from the Euclidean distances of the sample projections in the two-dimensional space.

3.2.2 Functional relatedness of module genes assessed by internal analysis

Based on the assumption that genes with similar patterns of expression across a vast variety of tissues are likely to encode proteins with related functions, we

selected from the literature three functional and well-characterized membrane trafficking complexes, and compared their adjacency distributions, and that of the MTM genes, with the adjacency scores of a null distribution, obtained as explained in section 2.3.1.3 (Fig. 3.12).

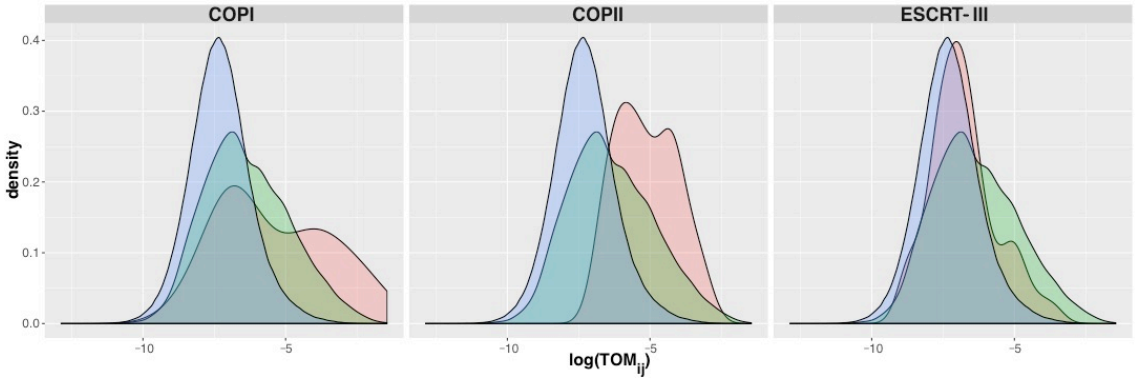


Figure 3.12 Density plots of topological overlap measures (TOMs) between pairs of trafficking genes forming different distributions. In violet the null distribution, in green the consensus distribution and in red the distribution of three membrane trafficking complexes, one per panel. TOMs are plotted on logarithmic scale.

The selected complexes were the coatomers of anterograde and retrograde vesicles (COPII and COPI complexes, respectively), and the machinery involved in MVB biogenesis for the selective degradation of cytoplasmic proteins (ESCRT-III complex); a summary of the gene products participating in the formation of the complexes is provided in Table 3.4.

As expected, the adjacency distributions of the complexes were significantly greater than the null distribution. Of note, the MTM distribution was also significantly greater than the null distribution, even more so than the COPI and ESCRT-III complexes, and at least as significant as that of the COPII complex (results in Table 3.5). This was the first indication that the module genes could be functionally related.

TABLE 3.4	
Gene	**Description**
COPI	
COPA	coatomer protein complex subunit alpha
COPB1	coatomer protein complex subunit beta 1
COPB2	coatomer protein complex subunit beta 2 (beta prime)
COPE	coatomer protein complex subunit epsilon
COPG1	coatomer protein complex subunit gamma 1
COPG2	coatomer protein complex subunit gamma 2
COPZ1	coatomer protein complex subunit zeta 1
COPZ2	coatomer protein complex subunit zeta 2
COPII	
PREB	prolactin regulatory element binding
SAR1A	secretion associated, Ras related GTPase 1A
SAR1B	secretion associated, Ras related GTPase 1B
SEC13	SEC13 homolog, nuclear pore and COPII coat complex component
SEC23A	Sec23 homolog A, COPII coat complex component
SEC23A	Sec23 homolog A, COPII coat complex component
SEC23B	Sec23 homolog B, COPII coat complex component
SEC24B	SEC24 homolog B, COPII coat complex component
SEC24C	SEC24 homolog C, COPII coat complex component
SEC24D	SEC24 homolog D, COPII coat complex component
SEC31A	SEC31 homolog A, COPII coat complex component
SEC31B	SEC31 homolog B, COPII coat complex component
ESCRT-III	
CHMP1A	charged multivesicular body protein 1A
CHMP1B	charged multivesicular body protein 1B
CHMP2A	charged multivesicular body protein 2A
CHMP2B	charged multivesicular body protein 2B
CHMP3	charged multivesicular body protein 3
CHMP4A	charged multivesicular body protein 4A
CHMP4B	charged multivesicular body protein 4B
CHMP4C	charged multivesicular body protein 4C
CHMP5	charged multivesicular body protein 5
CHMP6	charged multivesicular body protein 6
CHMP7	charged multivesicular body protein 7

Table 3.4 Summary of the gene products participating in the formation of the complexes.

TABLE 3.5	*x*			
H0:x≤null distr. **Ha:**x>null distr.	*consensus disr.*	*COPI distr.*	*COPII distr.*	*ESCRT-III distr.*
p.value	< 2.2e-16	1.5e-06	< 2.2e-16	0.4e-02

Table 3.5 Results of Wilcoxon rank-sum test (I). We tested the hypothesis of *x* less than or equal to the null distribution.

3.2.3 Functional relatedness of module genes confirmed by external sources

Functional association was reassessed using external data sources: the GIANT all-tissue network and the STRING database (see sections 2.3.2.1 and 2.3.2.2).

The GIANT all-tissue network stores connectivity scores for combinations of 25,825 genes, while the version 10.0 of the human STRING database provides scores for 3,853,379 pairs of proteins, associated with 17,946 official gene symbols, according to the annotation of the Ensembl (Hubbard et al., 2002) Archive Release 79 (http://mar2015.archive.ensembl.org).

For each source, we compared the interaction probabilities between the module gene pairs and the residual trafficking gene pairs (Fig. 3.13), under the assumption that the first distribution was lower than the second one (see section 2.3.2.3 for details). We restricted the data set to trafficking genes as we lacked information about the associations of the other genes.

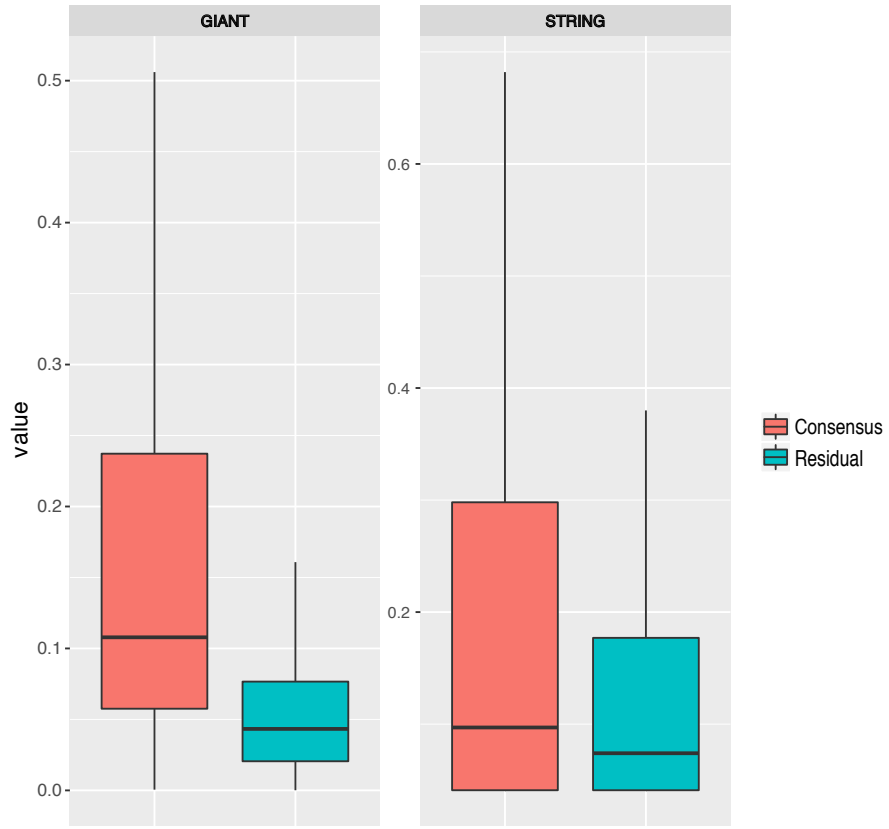


Figure 3.13 Box plots of interaction probabilities between the module gene pairs or the residual trafficking gene pairs. On the left, interaction probabilities from GIANT; on the right, those from STRING.

From the GIANT network we extracted 783,126 scores, corresponding to each pair of 1,252 trafficking genes; of these, 6,271 belonged to the MTM gene pairs. Only 58,294 cumulative scores, instead, were found between the protein products of 1,207 trafficking genes available in the STRING database, 879 of which were associated to the module genes.

In both cases the null hypothesis was rejected, supporting the thesis that these genes could functionally interact. It should be emphasized that the difference between the two distributions was more evident for GIANT than STRING (p-value<$2.2e^{-16}$ and p-value=$4.3e^{-14}$, respectively). This might be due either to a lack of evidence for most pairs of trafficking genes in STRING, or to the fact that STRING does not consider whether an interaction is preserved in multiple tissues, as GIANT does.

3.3 Consensus MTM transcriptional regulation

3.3.1 A regulatory network from TF binding site predictions

Groups of co-expressed genes are likely to be regulated, directly or indirectly, by common transcription factors. We identified putative TFs involved in the regulation of the expression levels of the module genes, employing a random walk with restart on a transcriptional regulatory network. To build the network (see section 2.4.1), we selected, from the Transcription Factor Target Gene Database, 360,387 significant DNA recognition motifs of 212 TFs, found in the proximal promoter region of 7,469 genes. These genes and TFs were selected applying the same criteria used for the selection of the consensus genes (detection in at least 22 out of 25 tissues).

3.3.2 Strategies employed to overcome the limits of random walks

Previous work has already demonstrated the power of random walks in predicting meaningful features of a given gene set (see section 1.4.4); biological networks that

combine multiple types of gene–gene or gene–property relationships are explored to identify genes mostly related to the set in the input.

However, these genes are identified from large and heterogeneous networks of biological information whose paths often converge onto hubs that connect different functional modules of the networks, and therefore not exclusively related to the set of genes in the input.

Additionally, it is possible that the gene you are looking for already resides in the set in the input but it is not detected because the probability of self-relatedness is altered by the probability of restarting from that node at each step of the random walk algorithm. A perfect example, in this regard, is the transcriptional regulatory network, which comprises interactions between regulators (transcription factors) and their target genes. It is a fact that some TFs are regulated by their own products, a phenomenon known as autogenous regulation, or by auto-regulation; thus, to allow the identification of these cases, we modified our regulatory network as described in section 2.4.2.

Staying on the topic of regulatory networks, we also demonstrated that correcting the random walk estimates, representing the likelihood of each TF in the network of regulating a precise set of genes, for those determined by rerunning the algorithm from different genes (background correction; see section 2.4.2), we significantly increased the probability of finding the right TF involved in the regulation of its targets in the top positions (Fig. 3.14 panel A). This outcome was not achieved by chance, as confirmed by the correlation of the probability vectors associated to each ChIP-seq experiment (see section 2.4.3; Fig. 3.14 panel B); the median correlation, in fact, was around the zero, and the highest coefficients were obtained by correlating the probability vectors associated to ChIP-seq experiments

covering either the same TF across different cell lines or different TFs belonging to the same family.

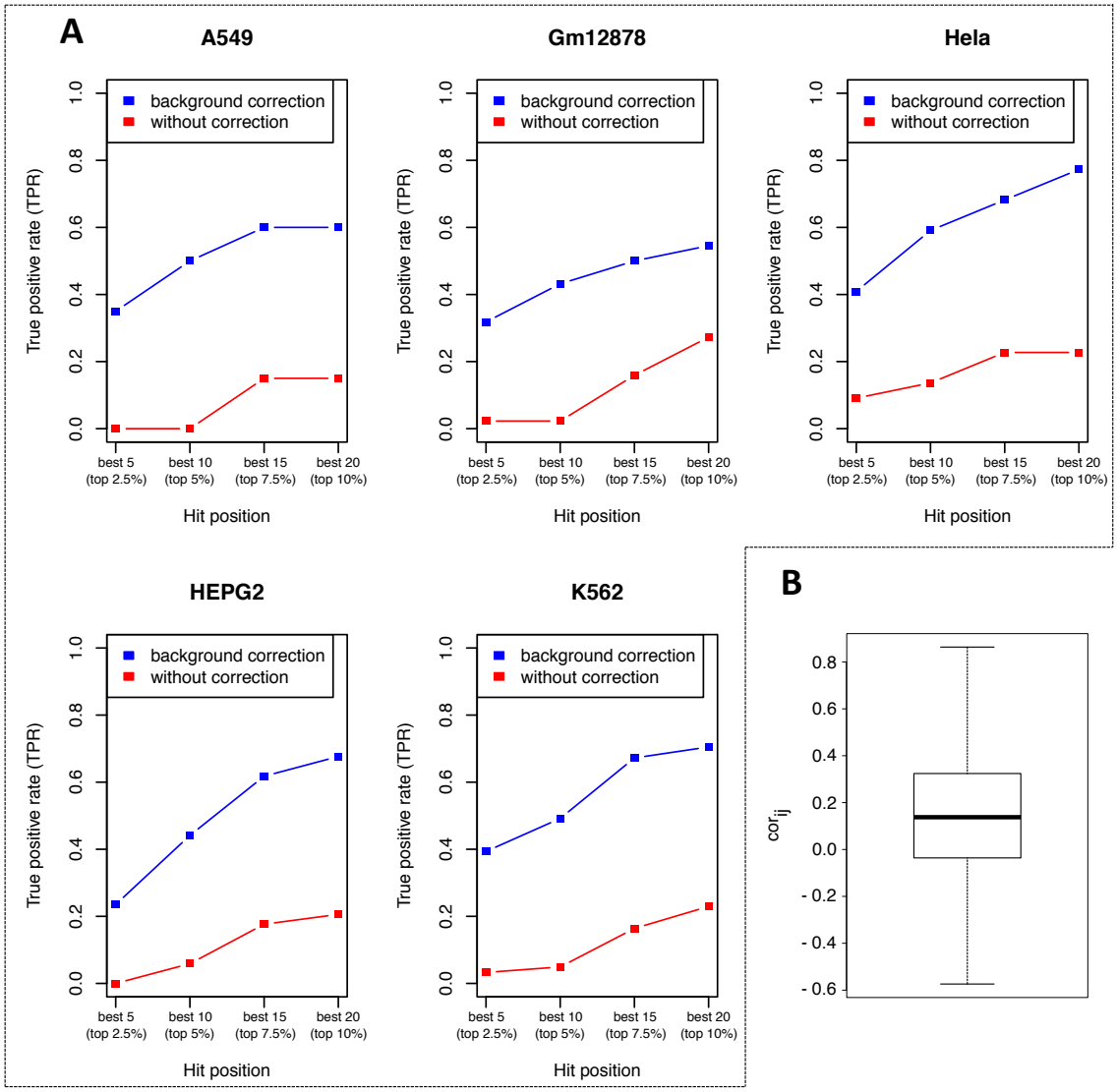


Figure 3.14 Evaluation of the background correction. A. Line plots showing the percentage of correct TFs ranked in the top positions (results are grouped by cell line). Rankings were obtained by restarting the random walk from the outcomes of ChIP-seq experiments. In red and in blue, respectively, percentages obtained before and after the correction. **B.** Correlation of the probability vectors associated to each ChIP-seq experiment.

3.3.3 Understanding the TFs controlling the expression of module genes

The random walk algorithm was then used to calculate the impact of each TF in the network over the regulation of the module genes. Estimates were corrected for the probabilities of the same TFs to regulate the genes from the remaining modules. The 10 best TFs of each consensus module are reported in supplementary Table 2.

3.4 Consensus MTM functional annotation

3.4.1 Putative role of MTM8 in collagen secretion

A putative biological role of the consensus MTM8 was found by chance analyzing the human cellular reprogramming of hiF-T cells to pluripotency. Clustering analysis of the RNA-Seq data from Cacchiarelli et al., 2015 (see section 2.5.1 for details) identified 9 major dynamic expression patterns of trafficking genes (Fig. 3.15).

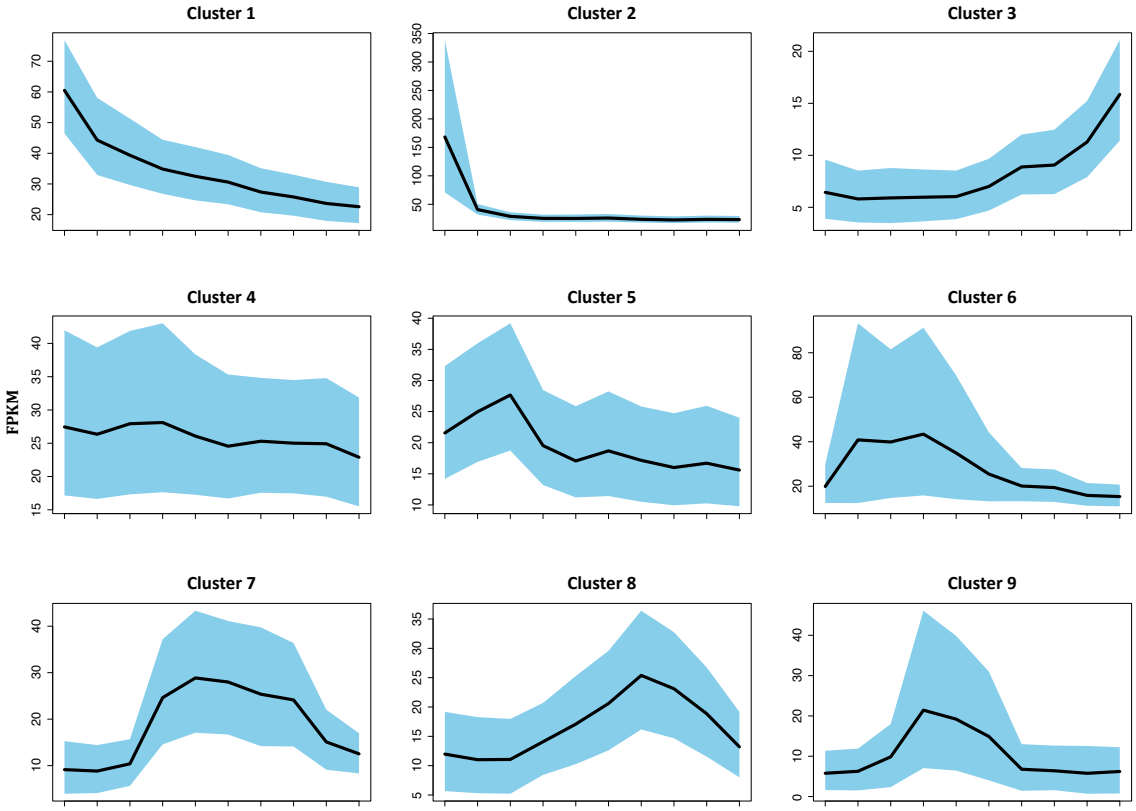


Figure 3.15 Line plots showing expression dynamics of differentially expressed trafficking genes during reprogramming. Cyan shades represent a 95% bootstrap confidence interval around the median values.

Specifically, cluster number 2 showed a trend similar to that of mesenchymal genes, including genes encoding structural components like collagen, that are rapidly down-regulated in the first stage of the reprogramming (Cacchiarelli et al., 2015). This observation was confirmed by annotation analyses reported in Figure 3.16 (panels A and B); in particular, the two complementary approaches used (see section 2.5.1) revealed an exclusive enrichment for both mesoderm-specific and somatic gene signatures.

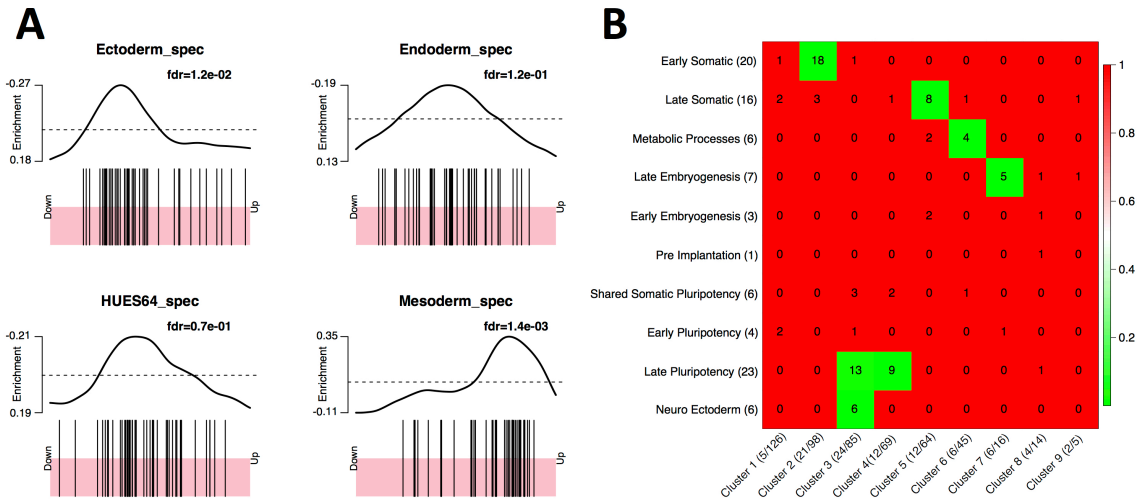


Figure 3.16 Characterization of cluster 2. A. Gene set enrichment analysis for condition-specific gene signatures obtained from human embryonic stem cells in their undifferentiated state (HUES64) or upon differentiation toward the three major germ layers (ectoderm, endoderm and mesoderm). **B.** Comparison with cluster of differentially expressed genes during reprogramming, annotated for developmental cell identity. Numbers in the table indicate gene counts in the intersection of the corresponding clusters; in brackets, the number of genes actually compared. Coloring of the table encodes the Fisher's exact test p-value for the overlap of the two clusters.

It is known that the mesoderm formed during gastrulation develops further into the paraxial mesoderm at early stages of embryogenesis, which later differentiates into somites; the somites in turn generate the mesenchymal connective tissue, to which fibroblasts belong (Gilbert, 2000). Thus, some of the trafficking genes in

cluster 2 could be part of the specific machinery involved in the secretion of collagen, the main protein synthesized and secreted by fibroblasts.

Since connective tissue comprises several tissues, including some GTEx categories (adipose tissue, blood, fibroblasts), and surrounds the majority of the other tissues, like brain and muscle, we tried to compare the trafficking gene clusters, identified during the reprogramming of hiF-T cells, to the consensus modules. Surprisingly, the consensus MTM8 showed a statistically significant (fdr<0.05) gene overlap with cluster 2; the genes in common between the two groups are reported in Table 3.6.

TABLE 3.6	
Gene	**Description**
ARF4	ADP ribosylation factor 4
CREB3L2	cAMP responsive element binding protein 3-like 2
DNAJC10	DnaJ heat shock protein family (Hsp40) member C10
KDELR2	KDEL endoplasmic reticulum protein retention receptor 2
PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2
SEC24D	SEC24 homolog D, COPII coat complex component
SEC31A	SEC31 homolog A, COPII coat complex component
SPCS3	signal peptidase complex subunit 3
SSR1	signal sequence receptor, alpha

Table 3.6 Summary of the genes in common between reprogramming cluster 2 and consensus MTM8.

Among these genes, there were many involved in different stages of the secretory pathway: translocation (SSR1), maturation (SPCS3), folding (DNAJC10) and export from the ER (SEC24D and SEC31A). There was also a transcription factor, CREB3L2, involved in the regulation of both SEC24D (Tomoishi et al., 2017) and ARF4 (Reiling et al., 2013), which was another gene on the list.

Furthermore, some of these genes were already reported to be required for collagen secretion, as in the case of SEC24D (Sarmah et al., 2010) and CREB3L2 (Ishikawa et al., 2017; Saito et al., 2009). Thus we hypothesized a possible role of the consensus MTM8 in collagen secretion.

3.5 Tissue-specific MTM detection and validation

3.5.1 A method to identify tissue-specific MTMs by differential co-expression analysis

To investigate if disease-related genes had specific membrane trafficking co-expressed communities in affected tissues, we identified tissue-specific MTMs combining differential correlation with weighted network analysis. Our method was not the first attempt to apply WGCNA to differential network analysis (Tesson et al., 2010), but it was the first one to explore differential co-expression patterns under different sample size conditions. Details of the methodology have been reported in section 2.6.1.

The accuracy of the method was evaluated by testing the differential co-expression of the synaptic vesicle genes in all the human tissues selected from GTEx (see section 2.6.1); the list of the synaptic vesicle genes examined at this stage is presented in Table 3.7.

TABLE 3.7	
Gene	**Description**
RAB3A	RAB3A, member RAS oncogene family
RAB3C	RAB3C, member RAS oncogene family
RAB5A	RAB5A, member RAS oncogene family
RAB7A	RAB7A, member RAS oncogene family
RPH3A	rabphilin 3A
SYN1	synapsin I
SYN2	synapsin II
SV2A	synaptic vesicle glycoprotein 2A
SYP	synaptophysin
SYT1	synaptotagmin 1
SYT2	synaptotagmin 2
SYT3	synaptotagmin 3
SYT5	synaptotagmin 5
SYT7	synaptotagmin 7
SYT8	synaptotagmin 8
SYT10	synaptotagmin 10
SYT11	synaptotagmin 11
SYT12	synaptotagmin 12
SYT13	synaptotagmin 13
SYT14	synaptotagmin 14
VAMP1	vesicle associated membrane protein 1
VAMP2	vesicle associated membrane protein 2
VAMP3	vesicle associated membrane protein 3
VAMP4	vesicle associated membrane protein 4
VAMP5	vesicle associated membrane protein 5

Table 3.7 Summary of the synaptic vesicle genes considered for the evaluation of the method.

There has been extensive literature on the co-expression of these genes in different brain areas: the four isoforms generated by alternative splicing from the transcripts of synapsin I and II (SYN1 and SYN2) genes, were reported to be co-expressed in almost all the brain regions (Südhof et al., 1989); many of the members of the synaptotagmin family were proven to be abundantly co-expressed in brain (Südhof, 2001); synaptotagmin I (SYT1) and synaptobrevin II (VAMP2) were shown to be co-expressed in the pineal gland of both rats and gerbils (Redecker, 1996).

In light of this evidence, we evaluated how accurate our method was in capturing the differential co-expression of the synaptic vesicle genes when using brain as reference tissue rather than the other tissues. Results shown in Figure 3.17 confirmed a significantly (p-value<$2.2e^{-16}$) high co-expression of these genes in brain compared to their co-expression in the other tissues.

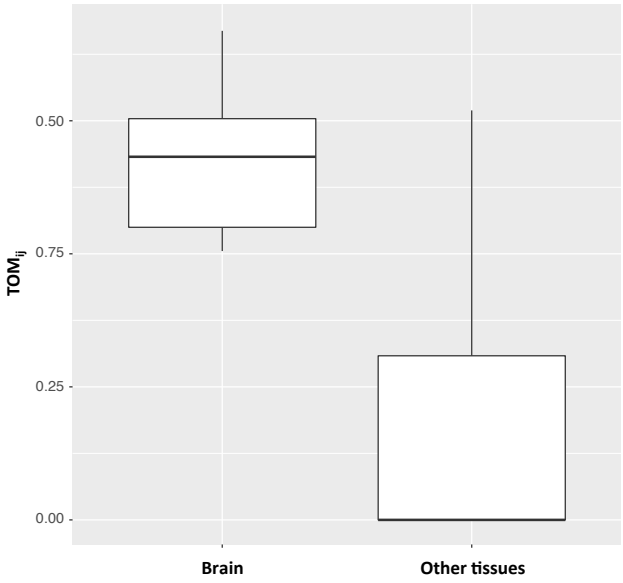


Figure 3.17 Box plots of topological overlap measures between the synaptic vesicle genes. In this particular case**,** TOMs were determined from the differences in the co-expression rates of the above-mentioned genes. On the left, the distribution of values obtained by using brain as reference tissue; on the right, the combined distribution obtained by using each of the other tissues as reference.

3.5.2 Validation of cutoff-survived modules

As for the consensus modules, tissue-specific MTMs were also defined by the cutreeDynamic function; in this case, function parameters were identified by evaluating, for each selected tissue, the number of modules produced and the average number of genes per module.

We found that, in general, the hybrid cut method produced less and larger modules than the tree cut method (Table 3.8-3.9). This was more evident when the hybrid cut method was used in combination with less sensitive splitting values; in a condition of high sensitivity to cluster splitting, instead, the number and the size of the modules were comparable between the two methods, except for some tissues for which the hybrid modules were still too large compared to those produced by the tree cut method that therefore was chosen to identify the tissue-specific MTMs.

TABLE 3.8	Hybrid (dS=0)	Hybrid (dS=1)	Hybrid (dS=2)	Hybrid (dS=3)	Hybrid (dS=4)	Dynamic Tree (No dS)	Dynamic Tree (dS)
Adipose Tissue	1	2	10	79	99	56	122
Adrenal Gland	2	5	28	83	92	49	115
Artery	1	1	4	34	68	32	127
Blood	1	1	1	26	50	25	102
Brain	1	1	1	18	60	42	114
Breast	1	1	2	14	51	23	113
Esophagus	1	1	1	1	2	11	83
Fibroblasts	1	1	1	17	49	30	100
Heart	1	1	6	38	76	42	118
Intestine	1	1	1	2	13	15	100
LCL	2	6	21	69	82	40	107
Liver	1	2	14	53	73	29	121
Lung	1	2	9	63	91	41	117
Muscle	1	2	10	68	90	47	109
Nerve	1	5	29	103	114	41	116
Ovary	7	19	65	103	106	57	124
Pancreas	1	3	42	95	104	65	126
Prostate	2	3	11	60	84	26	121
Reproductive Organs	3	6	27	80	90	48	116
Skin	1	1	11	68	93	46	131
Spleen	4	20	65	100	105	57	121
Stomach	1	1	1	3	11	13	79
Testis	1	2	3	35	61	28	119
Thyroid	1	5	25	94	104	52	127
Vagina	1	2	5	29	45	23	113

Table 3.8 Number of differentially co-expressed modules detected by different tree cut methods in each tissue. Values are colored from red (few modules) to green (many modules).

TABLE 3.9	Hybrid (dS=0)	Hybrid (dS=1)	Hybrid (dS=2)	Hybrid (dS=3)	Hybrid (dS=4)	Dynamic Tree (No dS)	Dynamic Tree (dS)
Adipose Tissue	1115	558	112	14	11	20	8
Adrenal Gland	563	225	40	14	12	23	8
Artery	1115	1115	279	33	16	35	8
Blood	1114	1114	1114	43	22	44	10
Brain	1147	1147	1147	64	19	27	9
Breast	1118	1118	559	80	22	48	9
Esophagus	1110	1110	1110	1110	555	100	12
Fibroblasts	1118	1118	1118	66	23	37	10
Heart	1117	1117	186	29	15	27	9
Intestine	1135	1135	1135	568	87	75	11
LCL	562	187	54	16	14	28	10
Liver	1126	563	80	21	15	38	9
Lung	1134	567	126	18	12	27	8
Muscle	1109	554	111	16	12	23	9
Nerve	1126	225	39	11	10	27	9
Ovary	161	59	17	11	11	19	8
Pancreas	1138	379	27	12	11	17	8
Prostate	570	380	104	19	14	44	8
Reproductive Organs	372	186	41	14	12	23	8
Skin	1130	1130	103	17	12	24	8
Spleen	283	57	17	11	11	20	9
Stomach	1117	1117	1117	372	102	85	13
Testis	1191	596	397	34	20	42	9
Thyroid	1137	227	45	12	11	21	8
Vagina	1127	564	225	39	25	49	9

Table 3.9 Average size of differentially co-expressed modules detected by different tree cut methods in each tissue. Values are colored from red (large modules) to green (small modules).

Internal validation was then used as a false positive handling process to resolve the issue of having too many differentially co-expressed modules. Here too, network-dependent and -independent preservation statistics were computed to statistically evaluate agreement of differential co-expression for cross-tissue comparisons (for details, see section 2.6.2; results are summarized in Figure 3.18). The differential co-expression of the internally validated modules for 19 human tissues was also evaluated in the relative GIANT interaction networks (see section 2.6.2); in 12 out of 19 cases, the interaction probability distribution of the module genes in the reference network was statistically different (fdr<0.05) from the combined distribution from all other tissues (results are shown in Figure 3.19 and Table 3.10). Since co-expression was the data type that had mostly contributed to

the GIANT scores between those genes, we concluded that the tissue-specific modules were confirmed to be differentially co-expressed also by GIANT networks.

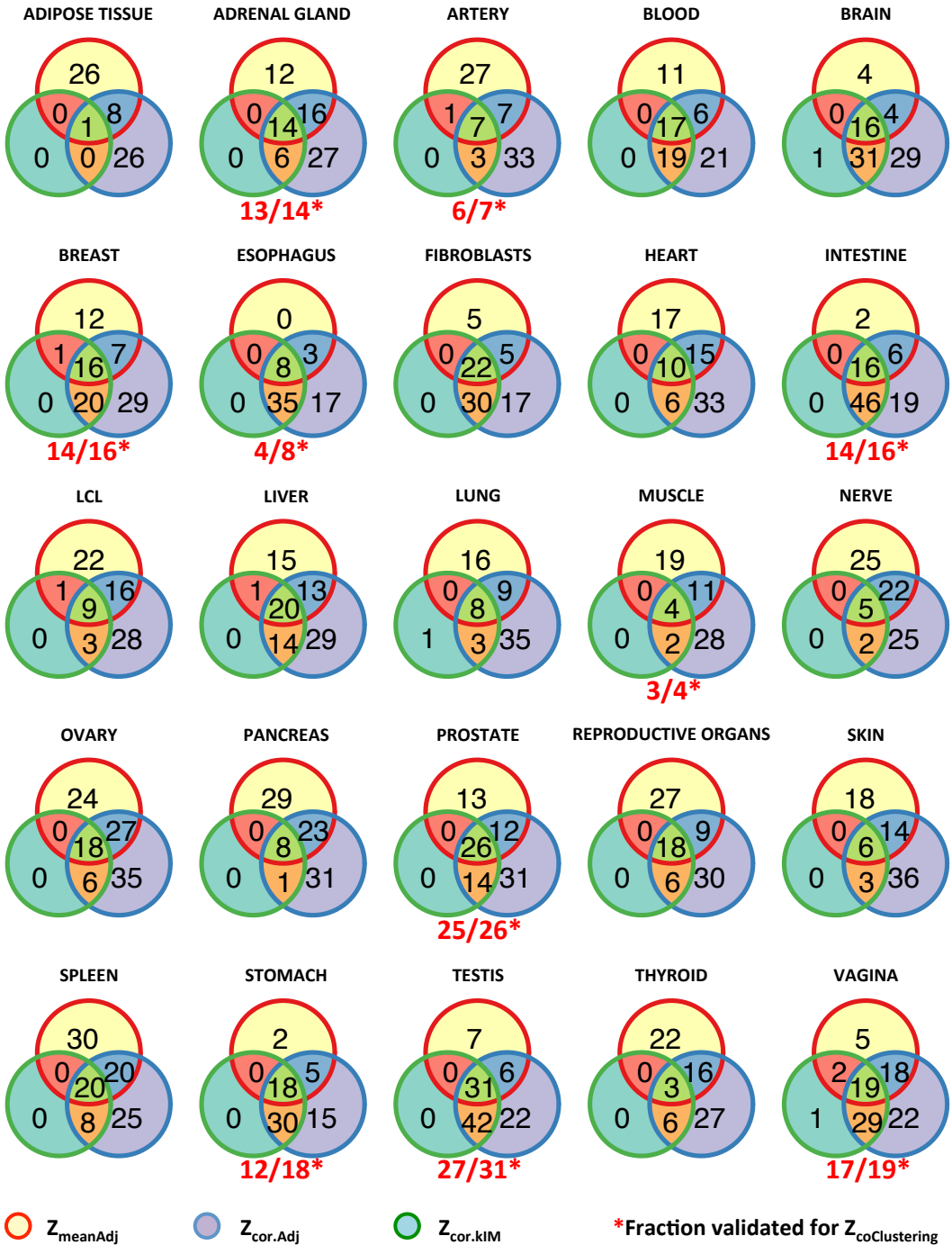


Figure 3.18 Venn diagrams showing the overlap of significantly preserved modules for three network-dependent statistics. In light green areas, the number of modules significantly preserved for all three statistics. Unless otherwise specified (red fractions at the bottom of the diagrams), the modules in the light green areas are also significantly preserved for the co-clustering preservation statistic.

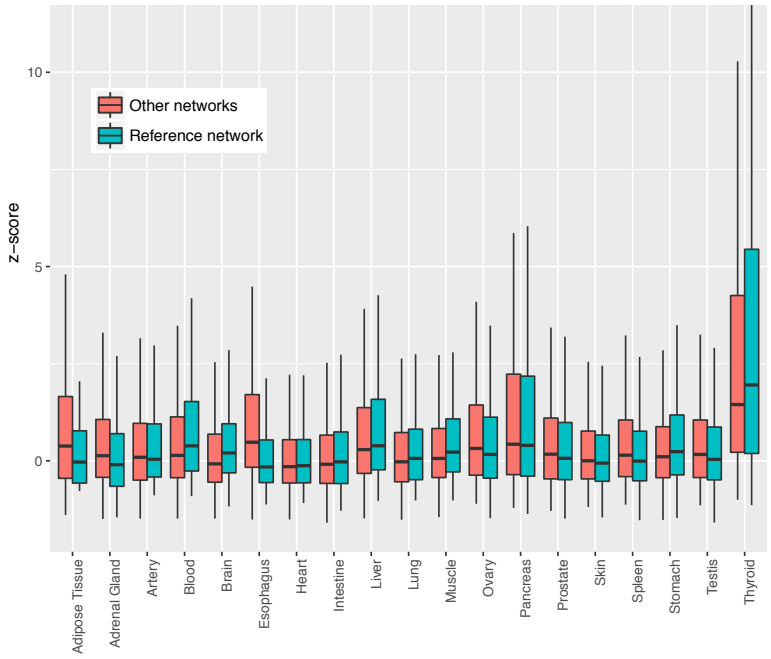


TABLE 3.10	**H0:**x=null distr. **Ha:**x≠null distr.
x	**fdr**
Adipose Tissue	2.6e-02
Adrenal Gland	1.0e-19
Artery	1.1e-01
Blood	5.2e-24
Brain	4.2e-21
Esophagus	1.3e-20
Heart	5.2e-01
Intestine	2.1e-01
Liver	3.4e-05
Lung	2.3e-03
Muscle	8.5e-04
Ovary	2.4e-04
Pancreas	9.4e-01
Prostate	5.0e-02
Skin	1.6e-01
Spleen	1.1e-08
Stomach	4.7e-08
Testis	4.8e-06
Thyroid	3.1e-01

Figure 3.19 Box plots of interaction probabilities between tissue-specific module genes in the reference tissue network or in the other tissue networks. Interaction probabilities were converted into Z-scores to make them comparable. **Table 3.10 Results of Wilcoxon rank-sum test (II)**. We tested the hypothesis of *x* equal to the null distribution (combined distribution of interaction probabilities between module genes from tissue networks other than reference network).

3.5.3 Examples of tissue-specific MTMs involving disease-related genes

As regards the disease-related genes, we chose two of them as a case study: TRAPPC2, mutated in the SpondyloEpiphysealDysplasia Tarda (SEDT; OMIM 313400), and ATP6AP2, muted in both X-linked mental retardation (MRXSH; OMIM 300556) and X-linked parkinsonism with spasticity (XPDS; OMIM 300911). Their modules were identified and validated for differential co-expression in the disease-relevant tissues, i.e. fibroblasts and brain, respectively, and the topology of these modules in the reference network versus the consensus network was plotted to highlight the differences in connection patterns (Figure 3.20 panels A and B).

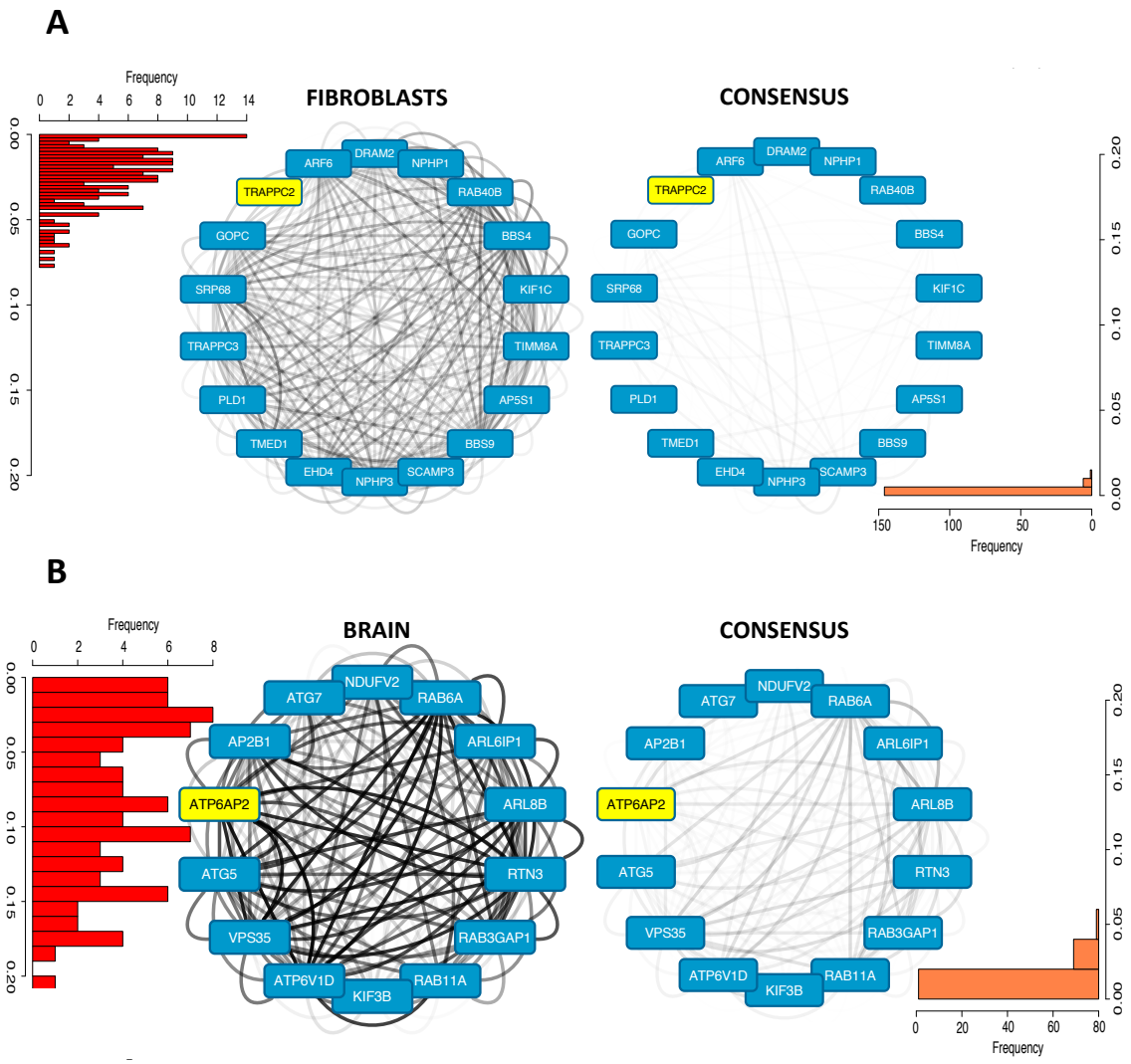


Figure 3.20 Example of modules differentially co-expressed across tissues. A-B Edges are colored according to their topological overlap with darker colors corresponding to higher topological overlap. The histograms close to each module represent the distribution of topological overlap of the network edges. The same set of genes is represented in the reference tissue network and in the consensus network. **A.** TRAPPC2 (highlighted in yellow) module in fibroblasts network versus consensus network. **B.** ATP6AP2 (highlighted in yellow) module in brain network versus consensus network.

Chapter 4

Discussion

4.1 Unidentified origin of certain consensuses

In the course of this project, we spent a lot of time identifying and characterizing consensus MTMs (paragraphs 3.1-3.4 refer to these aspects). A preliminary analysis, in fact, revealed that the majority of the trafficking genes were organized in co-expression modules, preserved among tissues (see section 3.1.5). These data were complemented by multidimensional scaling of tissue samples based on the expression of the trafficking genes (see section 3.2.1). Except for brain and blood samples that were clearly separated from the other tissues, the rest of the tissues shared high similarity in the traffikome signatures (Figure 3.10); this evidence suggested that only a small fraction of the genes could be differentially expressed, and thus co-expressed, across tissues.

However, not all the consensus modules identified showed significant similarity with the clusters of the preliminary analysis, although many of these modules were effectively preserved in several tissues having passed the internal validation (see section 3.2.1). We could only speculate that this consensus had become visible after the loss of tissue-specific co-expressions, during the computation of the consensus TOM (see section 2.2.4), which determined the segregation of those genes in different modules.

Not much can be said instead about the transcriptional regulation and function of these modules, at least until our predictions will be tested in the wet lab. In the near future, we would like to conduct a high-throughput RNA-Seq experiment to determine if the module genes respond to the silencing of the transcription factors, predicted to be involved in their regulation, in at least three different cell lines.

4.2 The unexploited power of random walks

In section 1.4.4 we discussed the application of the scoring properties of random walks to diverse biological issues. However, in section 3.3.2 we pointed out the limits of this stochastic process for detecting transcriptional regulatory mechanisms underlying sets of known TF-target genes and proposed a solution to improve the prediction of the correct TFs involved in the regulation of their targets. More generally, we feel that the potential of random walks has not being fully exploited, a proof of which comes from the analysis of the co-occurrence network of biomedical concepts provided by Grammatica et al., 2014 (see section 2.7.1).

During the analysis, we found that the amount of literature associated to different network entities (e.g. diseases) biased their random walk scores; in particular, highly studied entities such as common diseases (e.g. breast cancer is associated to 142,109 abstracts while hypeoxaluria to only 1,293), or with a longer experimental history, tended to be reached with higher random walk scores than other nodes in the network. To avoid this imbalance, a normalization procedure was applied to penalize entities reported in a relatively high number of manuscripts (see section 2.7.2). When applied on random walk scores connecting diseases to drugs, or drugs to other drugs, the normalization procedure allowed the linearization of the relationship between the total random walk sum and the reachability for each drug (Fig. 4.1 panel A), and improved the predictive performances of both networks (Fig. 4.1 panels B and C; for details on the benchmark visit section 2.7.3.1); specifically, it increased the power to infer drugs with similar mode of action (AUC RWs = 0.92 and AUC normalized RWs=0.96) and diseases sharing similar

mechanistic biological dysfunctions (AUC RWs~0.5, p-value=0.84 and AUC normalized RWs=0.65, p-value<$2.2e^{-16}$).

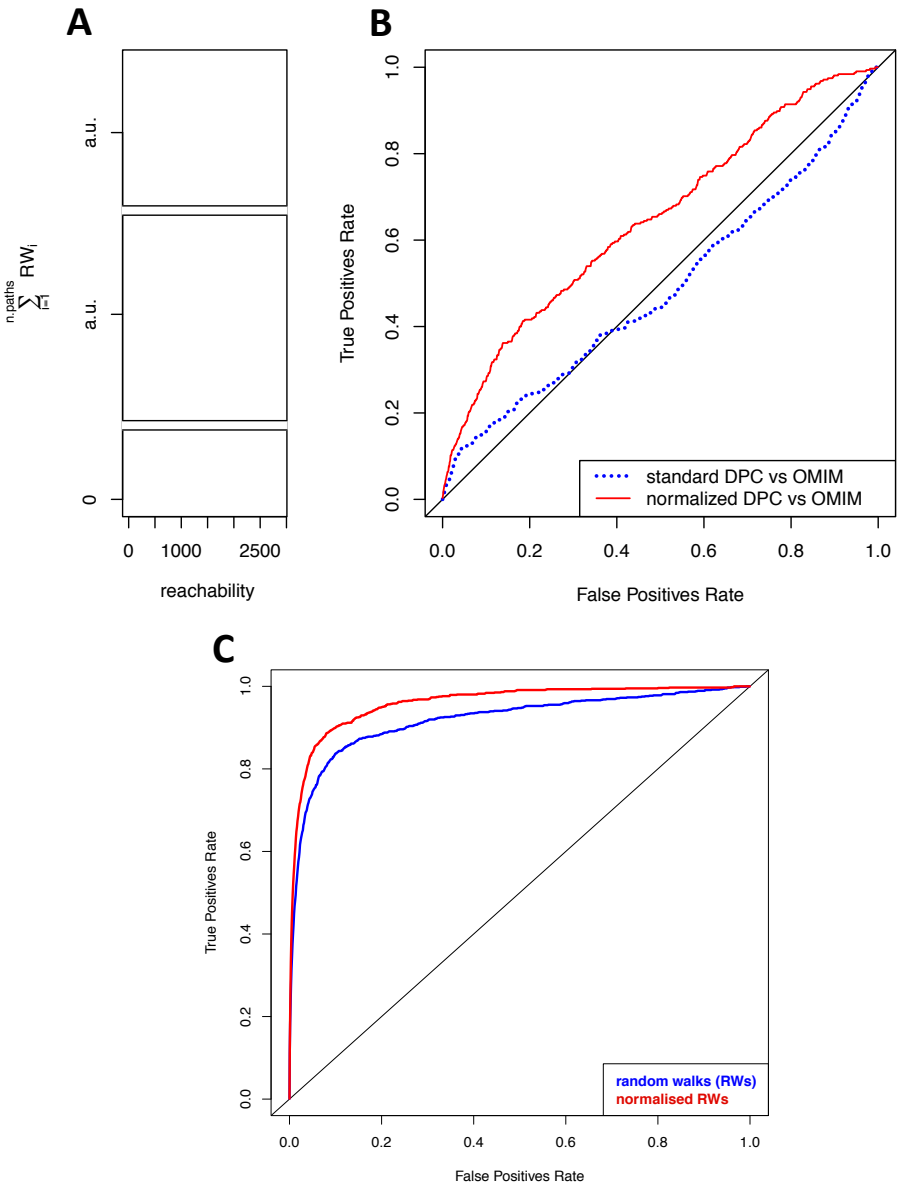


Figure 4.1 Assessing network predictive performances before and after score normalization. A Sum of random walk scores for each destination node of the drug-drug network, before (blue) and after (red) the normalization procedure. **B** Standard and normalized version of the disease-drug subnetwork against the OMIM-derived disease network. **C** Standard and normalized version of the drug-drug subnetwork against the ATC-derived drug network.

According to our findings, random walks are excellent tools for prioritizing nodes or edges of a network, but to be able to use their full potential the biological context in which the method is used should be considered, and, if necessary, the initial predictions should be corrected.

4.3 First steps towards the identification of tissue-specific MTMs and future applications

In sections 2.6.1 and 3.5.1 we presented a new method for the identification of tissue-specific MTMs. Although the methodology proved to be too sensitive to small variations between two or more conditions (since the fraction of internally validated modules was considerably lower than the predicted ones), external validation analysis showed that it was accurate in identifying differentially co-expressed genes (see section 3.5.2). Actually, we expected that the majority of the genes would not be assigned to any cluster, but in fact, just the opposite occurred (results are shown in Table 4.1).

TABLE 4.1	Hybrid (dS=0)	Hybrid (dS=1)	Hybrid (dS=2)	Hybrid (dS=3)	Hybrid (dS=4)	Dynamic Tree (No dS)	Dynamic Tree (dS)
Adipose Tissue	116	116	116	116	116	135	254
Adrenal Gland	105	105	105	105	105	124	257
Artery	116	116	116	116	116	120	205
Blood	117	117	117	117	117	131	224
Brain	84	84	84	84	84	97	211
Breast	113	113	113	113	113	119	228
Esophagus	121	121	121	121	121	133	255
Fibroblasts	113	113	113	113	113	123	235
Heart	114	114	114	114	114	117	222
Intestine	96	96	96	96	96	99	178
LCL	107	107	107	107	107	131	202
Liver	105	105	105	105	105	115	178
Lung	97	97	97	97	97	107	246
Muscle	122	122	122	122	122	144	250
Nerve	105	105	105	105	105	118	196
Ovary	107	107	107	107	107	130	222
Pancreas	93	93	93	93	93	115	248
Prostate	92	92	92	92	92	96	222
Reproductive Organs	114	114	114	114	114	134	258
Skin	101	101	101	101	101	116	213
Spleen	100	100	100	100	100	115	202
Stomach	114	114	114	114	114	121	237
Testis	40	40	40	40	40	43	124
Thyroid	94	94	94	94	94	116	199
Vagina	104	104	104	104	104	104	246

Table 4.1 Number of unassigned genes detected by different tree cut methods in each tissue. Values are colored from red (few genes) to green (many genes).

Our expectations came from the fact that several different tissues shared high similarities in their traffickome profiles (see paragraph 4.1), therefore only a small fraction of the genes was expected to account for tissue-specific interactions. To address these shortcomings the development of a new version of the method will focus on devising new strategies to make it less sensitive to small variations of gene co-expression between two or more conditions.

Regarding its application, the method has been used up to now to detect specific MTMs, involving disease-related genes, in those tissues that would exhibit the relevant phenotype. However, we did not consider membrane trafficking genes whose mutations caused multisystem disorders. The oculocerebrorenal syndrome of Lowe (OCRL) that primarily affects the eyes, nervous system, and kidneys represents a perfect example in this regard. This syndrome is caused by mutations in the OCRL1 gene, which encodes an inositol polyphosphate-5-phosphatase; although OCRL1 is ubiquitously expressed, the manifestations of Lowe syndrome are restricted only to the tissues mentioned above. In future work, we would like to deal with such cases, starting from OCRL, trying to combine the consensus analysis, which will be used to derive the oculocerebrorenal network or more generally multi-tissue networks, with differential co-expression analysis.

Appendix

Supplementary materials

SUPPLEMENTARY TABLES

Supplementary Table 1

Gene	Gene.description	kME
Cons.1		
APH1A	aph-1 homolog A, gamma-secretase subunit [Source:HGNC Symbol;Acc:HGNC:29509]	0.54
ARF5	ADP ribosylation factor 5 [Source:HGNC Symbol;Acc:HGNC:658]	0.76
ARFRP1	ADP ribosylation factor related protein 1 [Source:HGNC Symbol;Acc:HGNC:662]	0.46
ARL8A	ADP ribosylation factor like GTPase 8A [Source:HGNC Symbol;Acc:HGNC:25192]	0.48
ARSA	arylsulfatase A [Source:HGNC Symbol;Acc:HGNC:713]	0.42
ATG4D	autophagy related 4D cysteine peptidase [Source:HGNC Symbol;Acc:HGNC:20789]	0.59
ATP10D	ATPase phospholipid transporting 10D (putative) [Source:HGNC Symbol;Acc:HGNC:13549]	-0.50
ATP6V0E2	ATPase H+ transporting V0 subunit e2 [Source:HGNC Symbol;Acc:HGNC:21723]	0.28
ATP8A1	ATPase phospholipid transporting 8A1 [Source:HGNC Symbol;Acc:HGNC:13531]	-0.27
ATP9B	ATPase phospholipid transporting 9B (putative) [Source:HGNC Symbol;Acc:HGNC:13541]	0.17
ATPAF2	ATP synthase mitochondrial F1 complex assembly factor 2 [Source:HGNC Symbol;Acc:HGNC:18802]	0.45
BBS10	Bardet-Biedl syndrome 10 [Source:HGNC Symbol;Acc:HGNC:26291]	-0.37
CHMP1A	charged multivesicular body protein 1A [Source:HGNC Symbol;Acc:HGNC:8740]	0.75
CHN1	chimerin 1 [Source:HGNC Symbol;Acc:HGNC:1943]	-0.35
CLASP2	cytoplasmic linker associated protein 2 [Source:HGNC Symbol;Acc:HGNC:17078]	-0.43
CLTB	clathrin light chain B [Source:HGNC Symbol;Acc:HGNC:2091]	0.71
COPZ2	coatomer protein complex subunit zeta 2 [Source:HGNC Symbol;Acc:HGNC:19356]	-0.03
CTTN	cortactin [Source:HGNC Symbol;Acc:HGNC:3338]	0.37
CYTH2	cytohesin 2 [Source:HGNC Symbol;Acc:HGNC:9502]	0.42
DGAT2	diacylglycerol O-acyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:16940]	0.18
DNM2	dynamin 2 [Source:HGNC Symbol;Acc:HGNC:2974]	0.57
DYSF	dysferlin [Source:HGNC Symbol;Acc:HGNC:3097]	-0.08
EHD3	EH domain containing 3 [Source:HGNC Symbol;Acc:HGNC:3244]	-0.04
EMD	emerin [Source:HGNC Symbol;Acc:HGNC:3331]	0.35
EPN1	epsin 1 [Source:HGNC Symbol;Acc:HGNC:21604]	0.82
FKRP	fukutin related protein [Source:HGNC Symbol;Acc:HGNC:17997]	0.34
FLOT2	flotillin 2 [Source:HGNC Symbol;Acc:HGNC:3758]	0.20
GAA	glucosidase alpha, acid [Source:HGNC Symbol;Acc:HGNC:4065]	0.29
GIT1	GIT ArfGAP 1 [Source:HGNC Symbol;Acc:HGNC:4272]	0.35
GORAB	golgin, RAB6 interacting [Source:HGNC Symbol;Acc:HGNC:25676]	-0.31
HPS1	HPS1, biogenesis of lysosomal organelles complex 3 subunit 1	0.34

	[Source:HGNC Symbol;Acc:HGNC:5163]	
INPP4B	inositol polyphosphate-4-phosphatase type II B [Source:HGNC Symbol;Acc:HGNC:6075]	-0.30
INPP5K	inositol polyphosphate-5-phosphatase K [Source:HGNC Symbol;Acc:HGNC:33882]	0.28
KALRN	kalirin RhoGEF kinase [Source:HGNC Symbol;Acc:HGNC:4814]	0.09
KIFC3	kinesin family member C3 [Source:HGNC Symbol;Acc:HGNC:6326]	0.37
MYO1D	myosin ID [Source:HGNC Symbol;Acc:HGNC:7598]	0.17
NAGLU	N-acetyl-alpha-glucosaminidase [Source:HGNC Symbol;Acc:HGNC:7632]	0.51
NAPA	NSF attachment protein alpha [Source:HGNC Symbol;Acc:HGNC:7641]	0.65
ORMDL3	ORMDL sphingolipid biosynthesis regulator 3 [Source:HGNC Symbol;Acc:HGNC:16038]	0.50
OSBP2	oxysterol binding protein 2 [Source:HGNC Symbol;Acc:HGNC:8504]	0.14
PEX10	peroxisomal biogenesis factor 10 [Source:HGNC Symbol;Acc:HGNC:8851]	0.54
PEX14	peroxisomal biogenesis factor 14 [Source:HGNC Symbol;Acc:HGNC:8856]	0.59
PICK1	protein interacting with PRKCA 1 [Source:HGNC Symbol;Acc:HGNC:9394]	0.55
PITPNM1	phosphatidylinositol transfer protein membrane associated 1 [Source:HGNC Symbol;Acc:HGNC:9003]	0.29
RAB11B	RAB11B, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9761]	0.55
RAB1B	RAB1B, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:18370]	0.77
RAB30	RAB30, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9770]	-0.42
RAB40C	RAB40C, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:18285]	0.51
RHOBTB3	Rho related BTB domain containing 3 [Source:HGNC Symbol;Acc:HGNC:18757]	-0.39
RHOD	ras homolog family member D [Source:HGNC Symbol;Acc:HGNC:670]	0.49
RILP	Rab interacting lysosomal protein [Source:HGNC Symbol;Acc:HGNC:30266]	0.46
RIN1	Ras and Rab interactor 1 [Source:HGNC Symbol;Acc:HGNC:18749]	0.37
RTN2	reticulon 2 [Source:HGNC Symbol;Acc:HGNC:10468]	0.24
SCAP	SREBF chaperone [Source:HGNC Symbol;Acc:HGNC:30634]	0.65
SLC9A1	solute carrier family 9 member A1 [Source:HGNC Symbol;Acc:HGNC:11071]	0.33
SNX21	sorting nexin family member 21 [Source:HGNC Symbol;Acc:HGNC:16154]	0.32
SNX27	sorting nexin family member 27 [Source:HGNC Symbol;Acc:HGNC:20073]	-0.25
STX17	syntaxin 17 [Source:HGNC Symbol;Acc:HGNC:11432]	-0.40
TAPBP	TAP binding protein [Source:HGNC Symbol;Acc:HGNC:11566]	0.41
TBC1D4	TBC1 domain family member 4 [Source:HGNC Symbol;Acc:HGNC:19165]	-0.22
TMED1	transmembrane p24 trafficking protein 1 [Source:HGNC	0.47

	Symbol;Acc:HGNC:17291]	
TMEM67	transmembrane protein 67 [Source:HGNC Symbol;Acc:HGNC:28396]	-0.41
TRAPPC6A	trafficking protein particle complex 6A [Source:HGNC Symbol;Acc:HGNC:23069]	0.49
TRAPPC9	trafficking protein particle complex 9 [Source:HGNC Symbol;Acc:HGNC:30832]	0.42
VPS26B	VPS26, retromer complex component B [Source:HGNC Symbol;Acc:HGNC:28119]	0.26
VPS4A	vacuolar protein sorting 4 homolog A [Source:HGNC Symbol;Acc:HGNC:13488]	0.61
VPS51	VPS51, GARP complex subunit [Source:HGNC Symbol;Acc:HGNC:1172]	0.59
ZFPL1	zinc finger protein like 1 [Source:HGNC Symbol;Acc:HGNC:12868]	0.57
Cons.2*		
ACSL5	acyl-CoA synthetase long chain family member 5 [Source:HGNC Symbol;Acc:HGNC:16526]	-0.01
ALG12	ALG12, alpha-1,6-mannosyltransferase [Source:HGNC Symbol;Acc:HGNC:19358]	0.57
AP3D1	adaptor related protein complex 3 delta 1 subunit [Source:HGNC Symbol;Acc:HGNC:568]	0.61
ARF1	ADP ribosylation factor 1 [Source:HGNC Symbol;Acc:HGNC:652]	0.76
ARFIP2	ADP ribosylation factor interacting protein 2 [Source:HGNC Symbol;Acc:HGNC:17160]	0.66
ARPC1B	actin related protein 2/3 complex subunit 1B [Source:HGNC Symbol;Acc:HGNC:704]	0.64
ATP6V0B	ATPase H+ transporting V0 subunit b [Source:HGNC Symbol;Acc:HGNC:861]	0.68
B4GALT1	beta-1,4-galactosyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:924]	0.48
BACE2	beta-site APP-cleaving enzyme 2 [Source:HGNC Symbol;Acc:HGNC:934]	0.30
BET1L	Bet1 golgi vesicular membrane trafficking protein like [Source:HGNC Symbol;Acc:HGNC:19348]	0.36
BICD1	BICD cargo adaptor 1 [Source:HGNC Symbol;Acc:HGNC:1049]	-0.21
BIN1	bridging integrator 1 [Source:HGNC Symbol;Acc:HGNC:1052]	0.01
CC2D2A	coiled-coil and C2 domain containing 2A [Source:HGNC Symbol;Acc:HGNC:29253]	-0.38
CLN6	CLN6, transmembrane ER protein [Source:HGNC Symbol;Acc:HGNC:2077]	0.56
COPE	coatomer protein complex subunit epsilon [Source:HGNC Symbol;Acc:HGNC:2234]	0.82
COPZ1	coatomer protein complex subunit zeta 1 [Source:HGNC Symbol;Acc:HGNC:2243]	0.70
CREB3	cAMP responsive element binding protein 3 [Source:HGNC Symbol;Acc:HGNC:2347]	0.44
DNAJC1	DnaJ heat shock protein family (Hsp40) member C1 [Source:HGNC Symbol;Acc:HGNC:20090]	0.54
DNMBP	dynamin binding protein [Source:HGNC Symbol;Acc:HGNC:30373]	0.06
DTNBP1	dystrobrevin binding protein 1 [Source:HGNC Symbol;Acc:HGNC:17328]	-0.05

DYNC2H1	dynein cytoplasmic 2 heavy chain 1 [Source:HGNC Symbol;Acc:HGNC:2962]	-0.52
ERGIC1	endoplasmic reticulum-golgi intermediate compartment 1 [Source:HGNC Symbol;Acc:HGNC:29205]	0.46
EXTL2	exostosin like glycosyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:3516]	-0.41
FUT8	fucosyltransferase 8 [Source:HGNC Symbol;Acc:HGNC:4019]	-0.12
GALNT2	polypeptide N-acetylgalactosaminyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:4124]	0.53
GALNT3	polypeptide N-acetylgalactosaminyltransferase 3 [Source:HGNC Symbol;Acc:HGNC:4125]	0.07
GLA	galactosidase alpha [Source:HGNC Symbol;Acc:HGNC:4296]	0.46
GOLGA2	golgin A2 [Source:HGNC Symbol;Acc:HGNC:4425]	0.33
GOLGA3	golgin A3 [Source:HGNC Symbol;Acc:HGNC:4426]	0.45
GOSR2	golgi SNAP receptor complex member 2 [Source:HGNC Symbol;Acc:HGNC:4431]	0.47
HGSNAT	heparan-alpha-glucosaminide N-acetyltransferase [Source:HGNC Symbol;Acc:HGNC:26527]	-0.35
LMAN2	lectin, mannose binding 2 [Source:HGNC Symbol;Acc:HGNC:16986]	0.84
LMNA	lamin A/C [Source:HGNC Symbol;Acc:HGNC:6636]	0.34
MPV17	MPV17, mitochondrial inner membrane protein [Source:HGNC Symbol;Acc:HGNC:7224]	0.38
P4HB	prolyl 4-hydroxylase subunit beta [Source:HGNC Symbol;Acc:HGNC:8548]	0.80
PDIA5	protein disulfide isomerase family A member 5 [Source:HGNC Symbol;Acc:HGNC:24811]	0.42
PGS1	phosphatidylglycerophosphate synthase 1 [Source:HGNC Symbol;Acc:HGNC:30029]	0.29
PLOD1	procollagen-lysine,2-oxoglutarate 5-dioxygenase 1 [Source:HGNC Symbol;Acc:HGNC:9081]	0.47
PLOD3	procollagen-lysine,2-oxoglutarate 5-dioxygenase 3 [Source:HGNC Symbol;Acc:HGNC:9083]	0.58
PMM2	phosphomannomutase 2 [Source:HGNC Symbol;Acc:HGNC:9115]	0.71
PPIB	peptidylprolyl isomerase B [Source:HGNC Symbol;Acc:HGNC:9255]	0.81
PREB	prolactin regulatory element binding [Source:HGNC Symbol;Acc:HGNC:9356]	0.69
RABGAP1L	RAB GTPase activating protein 1 like [Source:HGNC Symbol;Acc:HGNC:24663]	-0.51
RER1	retention in endoplasmic reticulum sorting receptor 1 [Source:HGNC Symbol;Acc:HGNC:30309]	0.64
RIN2	Ras and Rab interactor 2 [Source:HGNC Symbol;Acc:HGNC:18750]	-0.20
SCAMP3	secretory carrier membrane protein 3 [Source:HGNC Symbol;Acc:HGNC:10565]	0.74
SEC13	SEC13 homolog, nuclear pore and COPII coat complex component [Source:HGNC Symbol;Acc:HGNC:10697]	0.84
SEC24C	SEC24 homolog C, COPII coat complex component [Source:HGNC Symbol;Acc:HGNC:10705]	0.53
SEC61B	Sec61 translocon beta subunit [Source:HGNC Symbol;Acc:HGNC:16993]	0.83
SERPINH1	serpin family H member 1 [Source:HGNC Symbol;Acc:HGNC:1546]	0.34
SLC35A2	solute carrier family 35 member A2 [Source:HGNC	0.79

	Symbol;Acc:HGNC:11022]	
SLC35D1	solute carrier family 35 member D1 [Source:HGNC Symbol;Acc:HGNC:20800]	-0.15
SLC50A1	solute carrier family 50 member 1 [Source:HGNC Symbol;Acc:HGNC:30657]	0.76
SNX33	sorting nexin 33 [Source:HGNC Symbol;Acc:HGNC:28468]	-0.15
SSR2	signal sequence receptor subunit 2 [Source:HGNC Symbol;Acc:HGNC:11324]	0.71
ST3GAL5	ST3 beta-galactoside alpha-2,3-sialyltransferase 5 [Source:HGNC Symbol;Acc:HGNC:10872]	-0.24
STX5	syntaxin 5 [Source:HGNC Symbol;Acc:HGNC:11440]	0.50
TMED3	transmembrane p24 trafficking protein 3 [Source:HGNC Symbol;Acc:HGNC:28889]	0.63
TMED9	transmembrane p24 trafficking protein 9 [Source:HGNC Symbol;Acc:HGNC:24878]	0.88
TMEM237	transmembrane protein 237 [Source:HGNC Symbol;Acc:HGNC:14432]	-0.39
TRAPPC4	trafficking protein particle complex 4 [Source:HGNC Symbol;Acc:HGNC:19943]	0.51
ULK2	unc-51 like autophagy activating kinase 2 [Source:HGNC Symbol;Acc:HGNC:13480]	-0.53
VMP1	vacuole membrane protein 1 [Source:HGNC Symbol;Acc:HGNC:29559]	0.27
VPS37C	VPS37C, ESCRT-I subunit [Source:HGNC Symbol;Acc:HGNC:26097]	0.47
YIF1A	Yip1 interacting factor homolog A, membrane trafficking protein [Source:HGNC Symbol;Acc:HGNC:16688]	0.76
ZFYVE20	#N/A	-0.46
Cons.3*		
ANKFY1	ankyrin repeat and FYVE domain containing 1 [Source:HGNC Symbol;Acc:HGNC:20763]	-0.04
ANO1	anoctamin 1 [Source:HGNC Symbol;Acc:HGNC:21625]	-0.21
AP2B1	adaptor related protein complex 2 beta 1 subunit [Source:HGNC Symbol;Acc:HGNC:563]	0.39
AP3B1	adaptor related protein complex 3 beta 1 subunit [Source:HGNC Symbol;Acc:HGNC:566]	0.84
AP3M2	adaptor related protein complex 3 mu 2 subunit [Source:HGNC Symbol;Acc:HGNC:570]	-0.17
AP3S2	adaptor related protein complex 3 sigma 2 subunit [Source:HGNC Symbol;Acc:HGNC:571]	0.44
ARHGEF2	Rho/Rac guanine nucleotide exchange factor 2 [Source:HGNC Symbol;Acc:HGNC:682]	-0.11
ARL2BP	ADP ribosylation factor like GTPase 2 binding protein [Source:HGNC Symbol;Acc:HGNC:17146]	0.51
ARSB	arylsulfatase B [Source:HGNC Symbol;Acc:HGNC:714]	0.46
ATG4A	autophagy related 4A cysteine peptidase [Source:HGNC Symbol;Acc:HGNC:16489]	0.62
ATG7	autophagy related 7 [Source:HGNC Symbol;Acc:HGNC:16935]	0.33
ATP6V1B2	ATPase H+ transporting V1 subunit B2 [Source:HGNC Symbol;Acc:HGNC:854]	0.49
BBS9	Bardet-Biedl syndrome 9 [Source:HGNC Symbol;Acc:HGNC:30000]	-0.31

BICD2	BICD cargo adaptor 2 [Source:HGNC Symbol;Acc:HGNC:17208]	0.20
BLZF1	basic leucine zipper nuclear factor 1 [Source:HGNC Symbol;Acc:HGNC:1065]	0.53
CLTC	clathrin heavy chain [Source:HGNC Symbol;Acc:HGNC:2092]	0.87
COG7	component of oligomeric golgi complex 7 [Source:HGNC Symbol;Acc:HGNC:18622]	-0.25
COPA	coatomer protein complex subunit alpha [Source:HGNC Symbol;Acc:HGNC:2230]	0.76
COPB1	coatomer protein complex subunit beta 1 [Source:HGNC Symbol;Acc:HGNC:2231]	0.83
COPB2	coatomer protein complex subunit beta 2 [Source:HGNC Symbol;Acc:HGNC:2232]	0.87
COPG2	coatomer protein complex subunit gamma 2 [Source:HGNC Symbol;Acc:HGNC:2237]	0.29
CPNE1	copine 1 [Source:HGNC Symbol;Acc:HGNC:2314]	-0.38
CREB3L4	cAMP responsive element binding protein 3 like 4 [Source:HGNC Symbol;Acc:HGNC:18854]	-0.31
CTSL	cathepsin L [Source:HGNC Symbol;Acc:HGNC:2537]	0.37
DERL1	derlin 1 [Source:HGNC Symbol;Acc:HGNC:28454]	0.67
DRAM2	DNA damage regulated autophagy modulator 2 [Source:HGNC Symbol;Acc:HGNC:28769]	0.13
ERAP1	endoplasmic reticulum aminopeptidase 1 [Source:HGNC Symbol;Acc:HGNC:18173]	0.48
ERGIC2	ERGIC and golgi 2 [Source:HGNC Symbol;Acc:HGNC:30208]	0.64
ERN1	endoplasmic reticulum to nucleus signaling 1 [Source:HGNC Symbol;Acc:HGNC:3449]	0.06
GOLGA1	golgin A1 [Source:HGNC Symbol;Acc:HGNC:4424]	-0.23
GOLGA5	golgin A5 [Source:HGNC Symbol;Acc:HGNC:4428]	0.72
GORASP2	golgi reassembly stacking protein 2 [Source:HGNC Symbol;Acc:HGNC:17500]	0.74
HPS4	HPS4, biogenesis of lysosomal organelles complex 3 subunit 2 [Source:HGNC Symbol;Acc:HGNC:15844]	-0.52
NPC1	NPC intracellular cholesterol transporter 1 [Source:HGNC Symbol;Acc:HGNC:7897]	0.34
PIP4K2C	phosphatidylinositol-5-phosphate 4-kinase type 2 gamma [Source:HGNC Symbol;Acc:HGNC:23786]	0.39
PLCG1	phospholipase C gamma 1 [Source:HGNC Symbol;Acc:HGNC:9065]	-0.48
PLD2	phospholipase D2 [Source:HGNC Symbol;Acc:HGNC:9068]	-0.30
PLEKHF1	pleckstrin homology and FYVE domain containing 1 [Source:HGNC Symbol;Acc:HGNC:20764]	-0.22
RAB11FIP1	RAB11 family interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:30265]	0.17
RAB1A	RAB1A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9758]	0.82
RAB3IP	RAB3A interacting protein [Source:HGNC Symbol;Acc:HGNC:16508]	0.09
RAB40B	RAB40B, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:18284]	-0.35
SAR1A	secretion associated Ras related GTPase 1A [Source:HGNC Symbol;Acc:HGNC:10534]	0.65
SCFD1	sec1 family domain containing 1 [Source:HGNC Symbol;Acc:HGNC:20726]	0.67

SEC22B	SEC22 homolog B, vesicle trafficking protein (gene/pseudogene) [Source:HGNC Symbol;Acc:HGNC:10700]	0.72
SEC24A	SEC24 homolog A, COPII coat complex component [Source:HGNC Symbol;Acc:HGNC:10703]	0.79
SGMS2	sphingomyelin synthase 2 [Source:HGNC Symbol;Acc:HGNC:28395]	0.45
SLC17A5	solute carrier family 17 member 5 [Source:HGNC Symbol;Acc:HGNC:10933]	0.48
SLC26A2	solute carrier family 26 member 2 [Source:HGNC Symbol;Acc:HGNC:10994]	0.37
SNX19	sorting nexin 19 [Source:HGNC Symbol;Acc:HGNC:21532]	0.49
STX18	syntaxin 18 [Source:HGNC Symbol;Acc:HGNC:15942]	0.44
TFEB	transcription factor EB [Source:HGNC Symbol;Acc:HGNC:11753]	-0.37
TFRC	transferrin receptor [Source:HGNC Symbol;Acc:HGNC:11763]	0.41
TSNARE1	t-SNARE domain containing 1 [Source:HGNC Symbol;Acc:HGNC:26437]	-0.60
USO1	USO1 vesicle transport factor [Source:HGNC Symbol;Acc:HGNC:30904]	0.75
WIPI1	WD repeat domain, phosphoinositide interacting 1 [Source:HGNC Symbol;Acc:HGNC:25471]	0.40
YIPF5	Yip1 domain family member 5 [Source:HGNC Symbol;Acc:HGNC:24877]	0.77
ZMPSTE24	zinc metallopeptidase STE24 [Source:HGNC Symbol;Acc:HGNC:12877]	0.66
Cons.4		
ANK3	ankyrin 3 [Source:HGNC Symbol;Acc:HGNC:494]	0.06
AP1B1	adaptor related protein complex 1 beta 1 subunit [Source:HGNC Symbol;Acc:HGNC:554]	0.56
AP5B1	adaptor related protein complex 5 beta 1 subunit [Source:HGNC Symbol;Acc:HGNC:25104]	0.54
ARL6IP6	ADP ribosylation factor like GTPase 6 interacting protein 6 [Source:HGNC Symbol;Acc:HGNC:24048]	-0.09
ATG13	autophagy related 13 [Source:HGNC Symbol;Acc:HGNC:29091]	0.12
ATP6V0A1	ATPase H+ transporting V0 subunit a1 [Source:HGNC Symbol;Acc:HGNC:865]	0.26
ATXN2	ataxin 2 [Source:HGNC Symbol;Acc:HGNC:10555]	-0.18
BBS4	Bardet-Biedl syndrome 4 [Source:HGNC Symbol;Acc:HGNC:969]	-0.35
CHST10	carbohydrate sulfotransferase 10 [Source:HGNC Symbol;Acc:HGNC:19650]	0.02
CLN8	CLN8, transmembrane ER and ERGIC protein [Source:HGNC Symbol;Acc:HGNC:2079]	0.26
EHD4	EH domain containing 4 [Source:HGNC Symbol;Acc:HGNC:3245]	0.18
ELOVL1	ELOVL fatty acid elongase 1 [Source:HGNC Symbol;Acc:HGNC:14418]	0.66
EXOC7	exocyst complex component 7 [Source:HGNC Symbol;Acc:HGNC:23214]	0.12
EXT1	exostosin glycosyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:3512]	-0.02
GOLM1	golgi membrane protein 1 [Source:HGNC Symbol;Acc:HGNC:15451]	-0.10
HTT	huntingtin [Source:HGNC Symbol;Acc:HGNC:4851]	0.28
IDS	iduronate 2-sulfatase [Source:HGNC Symbol;Acc:HGNC:5389]	-0.25
INPP5B	inositol polyphosphate-5-phosphatase B [Source:HGNC	-0.23

	Symbol;Acc:HGNC:6077]	
ITPR1	inositol 1,4,5-trisphosphate receptor type 1 [Source:HGNC Symbol;Acc:HGNC:6180]	-0.19
MBOAT2	membrane bound O-acyltransferase domain containing 2 [Source:HGNC Symbol;Acc:HGNC:25193]	0.01
MPDU1	mannose-P-dolichol utilization defect 1 [Source:HGNC Symbol;Acc:HGNC:7207]	0.64
MTMR14	myotubularin related protein 14 [Source:HGNC Symbol;Acc:HGNC:26190]	0.24
PEX5	peroxisomal biogenesis factor 5 [Source:HGNC Symbol;Acc:HGNC:9719]	0.27
PIK3C2B	phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2 beta [Source:HGNC Symbol;Acc:HGNC:8972]	0.21
PLCB1	phospholipase C beta 1 [Source:HGNC Symbol;Acc:HGNC:15917]	-0.41
PRKAA2	protein kinase AMP-activated catalytic subunit alpha 2 [Source:HGNC Symbol;Acc:HGNC:9377]	-0.31
PRKG1	protein kinase, cGMP-dependent, type I [Source:HGNC Symbol;Acc:HGNC:9414]	-0.25
RAB20	RAB20, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:18260]	0.16
RAB2B	RAB2B, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:20246]	-0.35
RABIF	RAB interacting factor [Source:HGNC Symbol;Acc:HGNC:9797]	0.23
RCAN2	regulator of calcineurin 2 [Source:HGNC Symbol;Acc:HGNC:3041]	-0.19
RPH3AL	rabphilin 3A like (without C2 domains) [Source:HGNC Symbol;Acc:HGNC:10296]	0.26
RRAGB	Ras related GTP binding B [Source:HGNC Symbol;Acc:HGNC:19901]	-0.32
SCAMP2	secretory carrier membrane protein 2 [Source:HGNC Symbol;Acc:HGNC:10564]	0.71
SLC35B2	solute carrier family 35 member B2 [Source:HGNC Symbol;Acc:HGNC:16872]	0.29
SLC35G2	solute carrier family 35 member G2 [Source:HGNC Symbol;Acc:HGNC:28480]	-0.23
SNX11	sorting nexin 11 [Source:HGNC Symbol;Acc:HGNC:14975]	0.46
ST3GAL1	ST3 beta-galactoside alpha-2,3-sialyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:10862]	0.31
STIM1	stromal interaction molecule 1 [Source:HGNC Symbol;Acc:HGNC:11386]	0.37
SYBU	syntabulin [Source:HGNC Symbol;Acc:HGNC:26011]	0.05
SYN2	synapsin II [Source:HGNC Symbol;Acc:HGNC:11495]	0.25
SYNE1	spectrin repeat containing nuclear envelope protein 1 [Source:HGNC Symbol;Acc:HGNC:17089]	-0.13
SYTL2	synaptotagmin like 2 [Source:HGNC Symbol;Acc:HGNC:15585]	-0.49
TBC1D10A	TBC1 domain family member 10A [Source:HGNC Symbol;Acc:HGNC:23609]	0.21
TBC1D14	TBC1 domain family member 14 [Source:HGNC Symbol;Acc:HGNC:29246]	0.41
TBC1D16	TBC1 domain family member 16 [Source:HGNC Symbol;Acc:HGNC:28356]	0.38
TP53	tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998]	0.28
UNC93B1	unc-93 homolog B1, TLR signaling regulator [Source:HGNC	0.27

	Symbol;Acc:HGNC:13481]	
VPS33B	VPS33B, late endosome and lysosome associated [Source:HGNC Symbol;Acc:HGNC:12712]	0.27
VPS53	VPS53, GARP complex subunit [Source:HGNC Symbol;Acc:HGNC:25608]	0.06
YKT6	YKT6 v-SNARE homolog [Source:HGNC Symbol;Acc:HGNC:16959]	0.52
ZFYVE1	zinc finger FYVE-type containing 1 [Source:HGNC Symbol;Acc:HGNC:13180]	0.27
ZFYVE21	zinc finger FYVE-type containing 21 [Source:HGNC Symbol;Acc:HGNC:20760]	0.07
Cons.5		
AGAP5	ArfGAP with GTPase domain, ankyrin repeat and PH domain 5 [Source:HGNC Symbol;Acc:HGNC:23467]	-0.44
ANKRD27	ankyrin repeat domain 27 [Source:HGNC Symbol;Acc:HGNC:25310]	0.38
AP3M1	adaptor related protein complex 3 mu 1 subunit [Source:HGNC Symbol;Acc:HGNC:569]	0.69
ARHGAP17	Rho GTPase activating protein 17 [Source:HGNC Symbol;Acc:HGNC:18239]	-0.10
ATG3	autophagy related 3 [Source:HGNC Symbol;Acc:HGNC:20962]	0.67
BBS5	Bardet-Biedl syndrome 5 [Source:HGNC Symbol;Acc:HGNC:970]	-0.38
BLOC1S2	biogenesis of lysosomal organelles complex 1 subunit 2 [Source:HGNC Symbol;Acc:HGNC:20984]	0.50
BMP1	bone morphogenetic protein 1 [Source:HGNC Symbol;Acc:HGNC:1067]	-0.25
CCZ1	CCZ1 homolog, vacuolar protein trafficking and biogenesis associated [Source:HGNC Symbol;Acc:HGNC:21691]	0.59
COG1	component of oligomeric golgi complex 1 [Source:HGNC Symbol;Acc:HGNC:6545]	-0.08
COX10	COX10, heme A:farnesyltransferase cytochrome c oxidase assembly factor [Source:HGNC Symbol;Acc:HGNC:2260]	0.56
CTSF	cathepsin F [Source:HGNC Symbol;Acc:HGNC:2531]	-0.41
CYTH1	cytohesin 1 [Source:HGNC Symbol;Acc:HGNC:9501]	-0.31
DNM1	dynamin 1 [Source:HGNC Symbol;Acc:HGNC:2972]	-0.33
DPM1	dolichyl-phosphate mannosyltransferase subunit 1, catalytic [Source:HGNC Symbol;Acc:HGNC:3005]	0.48
DRAM1	DNA damage regulated autophagy modulator 1 [Source:HGNC Symbol;Acc:HGNC:25645]	0.17
EXOC2	exocyst complex component 2 [Source:HGNC Symbol;Acc:HGNC:24968]	0.41
GDI2	GDP dissociation inhibitor 2 [Source:HGNC Symbol;Acc:HGNC:4227]	0.80
GIT2	GIT ArfGAP 2 [Source:HGNC Symbol;Acc:HGNC:4273]	0.07
GNPAT	glyceronephosphate O-acyltransferase [Source:HGNC Symbol;Acc:HGNC:4416]	0.55
GOLPH3L	golgi phosphoprotein 3 like [Source:HGNC Symbol;Acc:HGNC:24882]	0.40
IFT20	intraflagellar transport 20 [Source:HGNC Symbol;Acc:HGNC:30989]	-0.16
INPPL1	inositol polyphosphate phosphatase like 1 [Source:HGNC Symbol;Acc:HGNC:6080]	-0.49
KIF16B	kinesin family member 16B [Source:HGNC Symbol;Acc:HGNC:15869]	0.51
KIF3B	kinesin family member 3B [Source:HGNC Symbol;Acc:HGNC:6320]	0.33
LBR	lamin B receptor [Source:HGNC Symbol;Acc:HGNC:6518]	0.34

MINPP1	multiple inositol-polyphosphate phosphatase 1 [Source:HGNC Symbol;Acc:HGNC:7102]	0.40
MKKS	McKusick-Kaufman syndrome [Source:HGNC Symbol;Acc:HGNC:7108]	0.56
MYO18A	myosin XVIIIA [Source:HGNC Symbol;Acc:HGNC:31104]	0.00
MYO1B	myosin IB [Source:HGNC Symbol;Acc:HGNC:7596]	0.24
NDRG2	NDRG family member 2 [Source:HGNC Symbol;Acc:HGNC:14460]	-0.23
OBSL1	obscurin like 1 [Source:HGNC Symbol;Acc:HGNC:29092]	-0.49
PEX12	peroxisomal biogenesis factor 12 [Source:HGNC Symbol;Acc:HGNC:8854]	0.27
PI4KB	phosphatidylinositol 4-kinase beta [Source:HGNC Symbol;Acc:HGNC:8984]	-0.40
PIGZ	phosphatidylinositol glycan anchor biosynthesis class Z [Source:HGNC Symbol;Acc:HGNC:30596]	-0.30
PLSCR1	phospholipid scramblase 1 [Source:HGNC Symbol;Acc:HGNC:9092]	0.22
POMT1	protein O-mannosyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:9202]	-0.65
RAB10	RAB10, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9759]	0.80
RFTN1	raftlin, lipid raft linker 1 [Source:HGNC Symbol;Acc:HGNC:30278]	-0.11
RTN3	reticulon 3 [Source:HGNC Symbol;Acc:HGNC:10469]	0.60
SNX2	sorting nexin 2 [Source:HGNC Symbol;Acc:HGNC:11173]	0.65
SNX3	sorting nexin 3 [Source:HGNC Symbol;Acc:HGNC:11174]	0.61
SNX5	sorting nexin 5 [Source:HGNC Symbol;Acc:HGNC:14969]	0.39
SNX6	sorting nexin 6 [Source:HGNC Symbol;Acc:HGNC:14970]	0.68
SNX7	sorting nexin 7 [Source:HGNC Symbol;Acc:HGNC:14971]	0.45
STX6	syntaxin 6 [Source:HGNC Symbol;Acc:HGNC:11441]	0.32
SYNCRIP	synaptotagmin binding cytoplasmic RNA interacting protein [Source:HGNC Symbol;Acc:HGNC:16918]	0.72
TBC1D8	TBC1 domain family member 8 [Source:HGNC Symbol;Acc:HGNC:17791]	-0.21
TK2	thymidine kinase 2, mitochondrial [Source:HGNC Symbol;Acc:HGNC:11831]	0.00
TOR1B	torsin family 1 member B [Source:HGNC Symbol;Acc:HGNC:11995]	0.42
ULK1	unc-51 like autophagy activating kinase 1 [Source:HGNC Symbol;Acc:HGNC:12558]	-0.45
VAPB	VAMP associated protein B and C [Source:HGNC Symbol;Acc:HGNC:12649]	0.49
VPS29	VPS29, retromer complex component [Source:HGNC Symbol;Acc:HGNC:14340]	0.44
VPS35	VPS35, retromer complex component [Source:HGNC Symbol;Acc:HGNC:13487]	0.76
VTI1A	vesicle transport through interaction with t-SNAREs 1A [Source:HGNC Symbol;Acc:HGNC:17792]	0.48
Cons.6*		
AP1AR	adaptor related protein complex 1 associated regulatory protein [Source:HGNC Symbol;Acc:HGNC:28808]	0.60
AP3S1	adaptor related protein complex 3 sigma 1 subunit [Source:HGNC Symbol;Acc:HGNC:2013]	0.57
AP4S1	adaptor related protein complex 4 sigma 1 subunit [Source:HGNC	0.30

	Symbol;Acc:HGNC:575]	
ARL14EP	ADP ribosylation factor like GTPase 14 effector protein [Source:HGNC Symbol;Acc:HGNC:26798]	0.50
ARL15	ADP ribosylation factor like GTPase 15 [Source:HGNC Symbol;Acc:HGNC:25945]	0.30
ARL5A	ADP ribosylation factor like GTPase 5A [Source:HGNC Symbol;Acc:HGNC:696]	0.69
ARL6IP1	ADP ribosylation factor like GTPase 6 interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:697]	0.58
ARL8B	ADP ribosylation factor like GTPase 8B [Source:HGNC Symbol;Acc:HGNC:25564]	0.81
ARRB1	arrestin beta 1 [Source:HGNC Symbol;Acc:HGNC:711]	-0.08
ATG5	autophagy related 5 [Source:HGNC Symbol;Acc:HGNC:589]	0.71
ATL2	atlastin GTPase 2 [Source:HGNC Symbol;Acc:HGNC:24047]	0.56
ATP6V1C1	ATPase H+ transporting V1 subunit C1 [Source:HGNC Symbol;Acc:HGNC:856]	0.58
ATP6V1G1	ATPase H+ transporting V1 subunit G1 [Source:HGNC Symbol;Acc:HGNC:864]	0.56
BLOC1S6	biogenesis of lysosomal organelles complex 1 subunit 6 [Source:HGNC Symbol;Acc:HGNC:8549]	0.70
CCZ1B	CCZ1 homolog B, vacuolar protein trafficking and biogenesis associated [Source:HGNC Symbol;Acc:HGNC:21717]	0.59
CD164	CD164 molecule [Source:HGNC Symbol;Acc:HGNC:1632]	0.65
CHM	CHM, Rab escort protein 1 [Source:HGNC Symbol;Acc:HGNC:1940]	0.72
CHML	CHM like, Rab escort protein 2 [Source:HGNC Symbol;Acc:HGNC:1941]	0.39
CHMP1B	charged multivesicular body protein 1B [Source:HGNC Symbol;Acc:HGNC:24287]	0.45
CHMP2B	charged multivesicular body protein 2B [Source:HGNC Symbol;Acc:HGNC:24537]	0.84
CHMP5	charged multivesicular body protein 5 [Source:HGNC Symbol;Acc:HGNC:26942]	0.61
CLCN2	chloride voltage-gated channel 2 [Source:HGNC Symbol;Acc:HGNC:2020]	-0.43
DYM	dymeclin [Source:HGNC Symbol;Acc:HGNC:21317]	0.54
EMC2	ER membrane protein complex subunit 2 [Source:HGNC Symbol;Acc:HGNC:28963]	0.67
ERO1L	#N/A	0.42
ESYT2	extended synaptotagmin 2 [Source:HGNC Symbol;Acc:HGNC:22211]	0.42
GABARAPL2	GABA type A receptor associated protein like 2 [Source:HGNC Symbol;Acc:HGNC:13291]	0.51
GALNS	galactosamine (N-acetyl)-6-sulfatase [Source:HGNC Symbol;Acc:HGNC:4122]	-0.38
GOLGA7	golgin A7 [Source:HGNC Symbol;Acc:HGNC:24876]	0.69
NECAP1	NECAP endocytosis associated 1 [Source:HGNC Symbol;Acc:HGNC:24539]	0.40
NPHP4	nephrocystin 4 [Source:HGNC Symbol;Acc:HGNC:19104]	-0.27
PEX13	peroxisomal biogenesis factor 13 [Source:HGNC Symbol;Acc:HGNC:8855]	0.53
PEX2	peroxisomal biogenesis factor 2 [Source:HGNC Symbol;Acc:HGNC:9717]	0.50

PIGA	phosphatidylinositol glycan anchor biosynthesis class A [Source:HGNC Symbol;Acc:HGNC:8957]	0.37
PLEKHF2	pleckstrin homology and FYVE domain containing 2 [Source:HGNC Symbol;Acc:HGNC:20757]	0.49
PRKCI	protein kinase C iota [Source:HGNC Symbol;Acc:HGNC:9404]	0.61
RAB11A	RAB11A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9760]	0.66
RAB14	RAB14, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:16524]	0.79
RAB18	RAB18, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:14244]	0.87
RAB2A	RAB2A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9763]	0.79
RAB3GAP2	RAB3 GTPase activating non-catalytic protein subunit 2 [Source:HGNC Symbol;Acc:HGNC:17168]	0.71
RAB6A	RAB6A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9786]	0.82
RAB9A	RAB9A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9792]	0.49
RABGAP1	RAB GTPase activating protein 1 [Source:HGNC Symbol;Acc:HGNC:17155]	0.42
RRM2B	ribonucleotide reductase regulatory TP53 inducible subunit M2B [Source:HGNC Symbol;Acc:HGNC:17296]	0.59
SCAMP1	secretory carrier membrane protein 1 [Source:HGNC Symbol;Acc:HGNC:10563]	0.75
SGMS1	sphingomyelin synthase 1 [Source:HGNC Symbol;Acc:HGNC:29799]	0.56
SNAP23	synaptosome associated protein 23 [Source:HGNC Symbol;Acc:HGNC:11131]	0.42
SNX16	sorting nexin 16 [Source:HGNC Symbol;Acc:HGNC:14980]	0.47
SNX25	sorting nexin 25 [Source:HGNC Symbol;Acc:HGNC:21883]	0.49
SNX4	sorting nexin 4 [Source:HGNC Symbol;Acc:HGNC:11175]	0.62
TAPBPL	TAP binding protein like [Source:HGNC Symbol;Acc:HGNC:30683]	-0.30
VAMP7	vesicle associated membrane protein 7 [Source:HGNC Symbol;Acc:HGNC:11486]	0.79
VAPA	VAMP associated protein A [Source:HGNC Symbol;Acc:HGNC:12648]	0.69
VPS26A	VPS26, retromer complex component A [Source:HGNC Symbol;Acc:HGNC:12711]	0.84
VPS41	VPS41, HOPS complex subunit [Source:HGNC Symbol;Acc:HGNC:12713]	0.51
YIPF6	Yip1 domain family member 6 [Source:HGNC Symbol;Acc:HGNC:28304]	0.65
Cons.7		
AGA	aspartylglucosaminidase [Source:HGNC Symbol;Acc:HGNC:318]	0.58
AP5M1	adaptor related protein complex 5 mu 1 subunit [Source:HGNC Symbol;Acc:HGNC:20192]	0.72
ARL4A	ADP ribosylation factor like GTPase 4A [Source:HGNC Symbol;Acc:HGNC:695]	0.09
ARSD	arylsulfatase D [Source:HGNC Symbol;Acc:HGNC:717]	0.34
ASAH1	N-acylsphingosine amidohydrolase 1 [Source:HGNC Symbol;Acc:HGNC:735]	0.65

ATG4C	autophagy related 4C cysteine peptidase [Source:HGNC Symbol;Acc:HGNC:16040]	0.46
ATP6V1A	ATPase H+ transporting V1 subunit A [Source:HGNC Symbol;Acc:HGNC:851]	0.82
ATP8B1	ATPase phospholipid transporting 8B1 [Source:HGNC Symbol;Acc:HGNC:3706]	0.39
CD34	CD34 molecule [Source:HGNC Symbol;Acc:HGNC:1662]	-0.20
CLCN3	chloride voltage-gated channel 3 [Source:HGNC Symbol;Acc:HGNC:2021]	0.81
CLCN4	chloride voltage-gated channel 4 [Source:HGNC Symbol;Acc:HGNC:2022]	0.32
CLCN6	chloride voltage-gated channel 6 [Source:HGNC Symbol;Acc:HGNC:2024]	-0.40
COX15	COX15, cytochrome c oxidase assembly homolog [Source:HGNC Symbol;Acc:HGNC:2263]	0.56
ERLIN2	ER lipid raft associated 2 [Source:HGNC Symbol;Acc:HGNC:1356]	0.72
FGD5	FYVE, RhoGEF and PH domain containing 5 [Source:HGNC Symbol;Acc:HGNC:19117]	-0.26
FUCA1	alpha-L-fucosidase 1 [Source:HGNC Symbol;Acc:HGNC:4006]	0.47
GALC	galactosylceramidase [Source:HGNC Symbol;Acc:HGNC:4115]	0.37
GDI1	GDP dissociation inhibitor 1 [Source:HGNC Symbol;Acc:HGNC:4226]	-0.60
GLIS2	GLIS family zinc finger 2 [Source:HGNC Symbol;Acc:HGNC:29450]	-0.41
GNPTAB	N-acetylglucosamine-1-phosphate transferase alpha and beta subunits [Source:HGNC Symbol;Acc:HGNC:29670]	0.53
HSD17B4	hydroxysteroid 17-beta dehydrogenase 4 [Source:HGNC Symbol;Acc:HGNC:5213]	0.63
ITPR3	inositol 1,4,5-trisphosphate receptor type 3 [Source:HGNC Symbol;Acc:HGNC:6182]	-0.21
KIF21A	kinesin family member 21A [Source:HGNC Symbol;Acc:HGNC:19349]	0.42
KIFAP3	kinesin associated protein 3 [Source:HGNC Symbol;Acc:HGNC:17060]	0.39
LAMP2	lysosomal associated membrane protein 2 [Source:HGNC Symbol;Acc:HGNC:6501]	0.64
LRBA	LPS responsive beige-like anchor protein [Source:HGNC Symbol;Acc:HGNC:1742]	0.60
LRP1	LDL receptor related protein 1 [Source:HGNC Symbol;Acc:HGNC:6692]	-0.37
MIA3	MIA family member 3, ER export factor [Source:HGNC Symbol;Acc:HGNC:24008]	0.48
NSF	N-ethylmaleimide sensitive factor, vesicle fusing ATPase [Source:HGNC Symbol;Acc:HGNC:8016]	0.61
OSBPL5	oxysterol binding protein like 5 [Source:HGNC Symbol;Acc:HGNC:16392]	-0.46
PEX7	peroxisomal biogenesis factor 7 [Source:HGNC Symbol;Acc:HGNC:8860]	0.56
PIGM	phosphatidylinositol glycan anchor biosynthesis class M [Source:HGNC Symbol;Acc:HGNC:18858]	0.52
PITPNM2	phosphatidylinositol transfer protein membrane associated 2 [Source:HGNC Symbol;Acc:HGNC:21044]	-0.44
PLA2G4A	phospholipase A2 group IVA [Source:HGNC Symbol;Acc:HGNC:9035]	0.19
PLCD4	phospholipase C delta 4 [Source:HGNC Symbol;Acc:HGNC:9062]	-0.11

RAB11FIP3	RAB11 family interacting protein 3 [Source:HGNC Symbol;Acc:HGNC:17224]	-0.61
RAB4A	RAB4A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9781]	0.37
RRAGD	Ras related GTP binding D [Source:HGNC Symbol;Acc:HGNC:19903]	0.50
SGPL1	sphingosine-1-phosphate lyase 1 [Source:HGNC Symbol;Acc:HGNC:10817]	0.43
SNX18	sorting nexin 18 [Source:HGNC Symbol;Acc:HGNC:19245]	0.36
SNX24	sorting nexin 24 [Source:HGNC Symbol;Acc:HGNC:21533]	0.35
SNX30	sorting nexin family member 30 [Source:HGNC Symbol;Acc:HGNC:23685]	0.31
SRC	SRC proto-oncogene, non-receptor tyrosine kinase [Source:HGNC Symbol;Acc:HGNC:11283]	-0.45
TRAK1	trafficking kinesin protein 1 [Source:HGNC Symbol;Acc:HGNC:29947]	0.23
TRIM32	tripartite motif containing 32 [Source:HGNC Symbol;Acc:HGNC:16380]	0.15
TRIP10	thyroid hormone receptor interactor 10 [Source:HGNC Symbol;Acc:HGNC:12304]	-0.53
TTC8	tetratricopeptide repeat domain 8 [Source:HGNC Symbol;Acc:HGNC:20087]	0.34
VPS45	vacuolar protein sorting 45 homolog [Source:HGNC Symbol;Acc:HGNC:14579]	0.23
Cons.8*		
ARF4	ADP ribosylation factor 4 [Source:HGNC Symbol;Acc:HGNC:655]	0.89
ARFGAP3	ADP ribosylation factor GTPase activating protein 3 [Source:HGNC Symbol;Acc:HGNC:661]	0.63
ARL1	ADP ribosylation factor like GTPase 1 [Source:HGNC Symbol;Acc:HGNC:692]	0.74
BET1	Bet1 golgi vesicular membrane trafficking protein [Source:HGNC Symbol;Acc:HGNC:14562]	0.40
CALU	calumenin [Source:HGNC Symbol;Acc:HGNC:1458]	0.73
CANX	calnexin [Source:HGNC Symbol;Acc:HGNC:1473]	0.77
CLCC1	chloride channel CLIC like 1 [Source:HGNC Symbol;Acc:HGNC:29675]	0.37
CLCN5	chloride voltage-gated channel 5 [Source:HGNC Symbol;Acc:HGNC:2023]	0.28
CREB3L2	cAMP responsive element binding protein 3 like 2 [Source:HGNC Symbol;Acc:HGNC:23720]	0.45
DERL2	derlin 2 [Source:HGNC Symbol;Acc:HGNC:17943]	0.63
DNAJB9	DnaJ heat shock protein family (Hsp40) member B9 [Source:HGNC Symbol;Acc:HGNC:6968]	0.73
DNAJC10	DnaJ heat shock protein family (Hsp40) member C10 [Source:HGNC Symbol;Acc:HGNC:24637]	0.57
EIF2AK3	eukaryotic translation initiation factor 2 alpha kinase 3 [Source:HGNC Symbol;Acc:HGNC:3255]	0.67
ELOVL5	ELOVL fatty acid elongase 5 [Source:HGNC Symbol;Acc:HGNC:21308]	0.48
ELOVL6	ELOVL fatty acid elongase 6 [Source:HGNC Symbol;Acc:HGNC:15829]	0.31
ERP44	endoplasmic reticulum protein 44 [Source:HGNC Symbol;Acc:HGNC:18311]	0.81
EXOC3	exocyst complex component 3 [Source:HGNC Symbol;Acc:HGNC:30378]	-0.50

GAS8	growth arrest specific 8 [Source:HGNC Symbol;Acc:HGNC:4166]	-0.24
GNPNAT1	glucosamine-phosphate N-acetyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:19980]	0.63
IQSEC1	IQ motif and Sec7 domain 1 [Source:HGNC Symbol;Acc:HGNC:29112]	-0.41
KDELR2	KDEL endoplasmic reticulum protein retention receptor 2 [Source:HGNC Symbol;Acc:HGNC:6305]	0.85
LMAN1	lectin, mannose binding 1 [Source:HGNC Symbol;Acc:HGNC:6631]	0.82
MCFD2	multiple coagulation factor deficiency 2 [Source:HGNC Symbol;Acc:HGNC:18451]	0.70
MGAT2	mannosyl (alpha-1,6-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase [Source:HGNC Symbol;Acc:HGNC:7045]	0.71
P4HA1	prolyl 4-hydroxylase subunit alpha 1 [Source:HGNC Symbol;Acc:HGNC:8546]	0.31
PACS1	phosphofurin acidic cluster sorting protein 1 [Source:HGNC Symbol;Acc:HGNC:30032]	-0.15
PAPSS2	3'-phosphoadenosine 5'-phosphosulfate synthase 2 [Source:HGNC Symbol;Acc:HGNC:8604]	0.28
PIK3CD	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta [Source:HGNC Symbol;Acc:HGNC:8977]	-0.30
PLOD2	procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 [Source:HGNC Symbol;Acc:HGNC:9082]	0.42
RAB27A	RAB27A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9766]	0.50
RPGRIP1L	RPGRIP1 like [Source:HGNC Symbol;Acc:HGNC:29168]	0.23
SEC23A	Sec23 homolog A, coat complex II component [Source:HGNC Symbol;Acc:HGNC:10701]	0.65
SEC24D	SEC24 homolog D, COPII coat complex component [Source:HGNC Symbol;Acc:HGNC:10706]	0.76
SEC31A	SEC31 homolog A, COPII coat complex component [Source:HGNC Symbol;Acc:HGNC:17052]	0.73
SERP1	stress associated endoplasmic reticulum protein 1 [Source:HGNC Symbol;Acc:HGNC:10759]	0.73
SLC35B4	solute carrier family 35 member B4 [Source:HGNC Symbol;Acc:HGNC:20584]	0.34
SPCS3	signal peptidase complex subunit 3 [Source:HGNC Symbol;Acc:HGNC:26212]	0.79
SSR1	signal sequence receptor subunit 1 [Source:HGNC Symbol;Acc:HGNC:11323]	0.87
STX1B	syntaxin 1B [Source:HGNC Symbol;Acc:HGNC:18539]	-0.45
TBC1D17	TBC1 domain family member 17 [Source:HGNC Symbol;Acc:HGNC:25699]	-0.66
TMED10	transmembrane p24 trafficking protein 10 [Source:HGNC Symbol;Acc:HGNC:16998]	0.80
TMED2	transmembrane p24 trafficking protein 2 [Source:HGNC Symbol;Acc:HGNC:16996]	0.92
TMED5	transmembrane p24 trafficking protein 5 [Source:HGNC Symbol;Acc:HGNC:24251]	0.70
VMA21	VMA21, vacuolar ATPase assembly factor [Source:HGNC Symbol;Acc:HGNC:22082]	0.54
VPS11	VPS11, CORVET/HOPS core subunit [Source:HGNC	-0.35

	Symbol;Acc:HGNC:14583]	
Cons.9*		
AP1S1	adaptor related protein complex 1 sigma 1 subunit [Source:HGNC Symbol;Acc:HGNC:559]	0.72
AP2S1	adaptor related protein complex 2 sigma 1 subunit [Source:HGNC Symbol;Acc:HGNC:565]	0.80
AP5S1	adaptor related protein complex 5 sigma 1 subunit [Source:HGNC Symbol;Acc:HGNC:15875]	0.63
ARL10	ADP ribosylation factor like GTPase 10 [Source:HGNC Symbol;Acc:HGNC:22042]	-0.50
ARL3	ADP ribosylation factor like GTPase 3 [Source:HGNC Symbol;Acc:HGNC:694]	0.24
ARPC4	actin related protein 2/3 complex subunit 4 [Source:HGNC Symbol;Acc:HGNC:707]	0.79
ATP6V0D1	ATPase H+ transporting V0 subunit d1 [Source:HGNC Symbol;Acc:HGNC:13724]	0.76
BLOC1S4	biogenesis of lysosomal organelles complex 1 subunit 4 [Source:HGNC Symbol;Acc:HGNC:24206]	0.43
CHMP2A	charged multivesicular body protein 2A [Source:HGNC Symbol;Acc:HGNC:30216]	0.77
CHMP6	charged multivesicular body protein 6 [Source:HGNC Symbol;Acc:HGNC:25675]	0.65
CTSZ	cathepsin Z [Source:HGNC Symbol;Acc:HGNC:2547]	0.39
EHBP1	EH domain binding protein 1 [Source:HGNC Symbol;Acc:HGNC:29144]	-0.37
ETHE1	ETHE1, persulfide dioxygenase [Source:HGNC Symbol;Acc:HGNC:23287]	0.45
FXN	frataxin [Source:HGNC Symbol;Acc:HGNC:3951]	0.33
INPP4A	inositol polyphosphate-4-phosphatase type I A [Source:HGNC Symbol;Acc:HGNC:6074]	-0.39
INVS	inversin [Source:HGNC Symbol;Acc:HGNC:17870]	-0.43
LAMTOR1	late endosomal/lysosomal adaptor, MAPK and MTOR activator 1 [Source:HGNC Symbol;Acc:HGNC:26068]	0.85
LRP6	LDL receptor related protein 6 [Source:HGNC Symbol;Acc:HGNC:6698]	-0.48
MPI	mannose phosphate isomerase [Source:HGNC Symbol;Acc:HGNC:7216]	0.43
NDUFS2	NADH:ubiquinone oxidoreductase core subunit S2 [Source:HGNC Symbol;Acc:HGNC:7708]	0.49
NDUFS3	NADH:ubiquinone oxidoreductase core subunit S3 [Source:HGNC Symbol;Acc:HGNC:7710]	0.74
OSBPL10	oxysterol binding protein like 10 [Source:HGNC Symbol;Acc:HGNC:16395]	-0.27
PGAP1	post-GPI attachment to proteins 1 [Source:HGNC Symbol;Acc:HGNC:25712]	-0.38
PLEKHA8	pleckstrin homology domain containing A8 [Source:HGNC Symbol;Acc:HGNC:30037]	-0.47
RAB13	RAB13, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9762]	0.35
RAB33B	RAB33B, member RAS oncogene family [Source:HGNC	-0.52

	Symbol;Acc:HGNC:16075]	
SCO2	SCO2, cytochrome c oxidase assembly protein [Source:HGNC Symbol;Acc:HGNC:10604]	0.39
SLC36A1	solute carrier family 36 member 1 [Source:HGNC Symbol;Acc:HGNC:18761]	0.25
SNF8	SNF8, ESCRT-II complex subunit [Source:HGNC Symbol;Acc:HGNC:17028]	0.76
SNX12	sorting nexin 12 [Source:HGNC Symbol;Acc:HGNC:14976]	0.50
SNX17	sorting nexin 17 [Source:HGNC Symbol;Acc:HGNC:14979]	0.71
STX2	syntaxin 2 [Source:HGNC Symbol;Acc:HGNC:3403]	-0.36
STX8	syntaxin 8 [Source:HGNC Symbol;Acc:HGNC:11443]	0.27
STXBP2	syntaxin binding protein 2 [Source:HGNC Symbol;Acc:HGNC:11445]	0.59
SURF1	SURF1, cytochrome c oxidase assembly factor [Source:HGNC Symbol;Acc:HGNC:11474]	0.52
TRAPPC1	trafficking protein particle complex 1 [Source:HGNC Symbol;Acc:HGNC:19894]	0.79
TRAPPC3	trafficking protein particle complex 3 [Source:HGNC Symbol;Acc:HGNC:19942]	0.45
VAMP8	vesicle associated membrane protein 8 [Source:HGNC Symbol;Acc:HGNC:12647]	0.62
VPS25	vacuolar protein sorting 25 homolog [Source:HGNC Symbol;Acc:HGNC:28122]	0.78
VPS72	vacuolar protein sorting 72 homolog [Source:HGNC Symbol;Acc:HGNC:11644]	0.27
VTI1B	vesicle transport through interaction with t-SNAREs 1B [Source:HGNC Symbol;Acc:HGNC:17793]	0.46
Cons.10		
ACTR1A	ARP1 actin related protein 1 homolog A [Source:HGNC Symbol;Acc:HGNC:167]	0.80
AP1M1	adaptor related protein complex 1 mu 1 subunit [Source:HGNC Symbol;Acc:HGNC:13667]	0.72
AP2A1	adaptor related protein complex 2 alpha 1 subunit [Source:HGNC Symbol;Acc:HGNC:561]	0.74
AP2M1	adaptor related protein complex 2 mu 1 subunit [Source:HGNC Symbol;Acc:HGNC:564]	0.67
ARAP2	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 2 [Source:HGNC Symbol;Acc:HGNC:16924]	-0.44
ATG2A	autophagy related 2A [Source:HGNC Symbol;Acc:HGNC:29028]	0.35
ATG9A	autophagy related 9A [Source:HGNC Symbol;Acc:HGNC:22408]	0.60
CHMP4B	charged multivesicular body protein 4B [Source:HGNC Symbol;Acc:HGNC:16171]	0.47
CHMP7	charged multivesicular body protein 7 [Source:HGNC Symbol;Acc:HGNC:28439]	0.43
CHST3	carbohydrate sulfotransferase 3 [Source:HGNC Symbol;Acc:HGNC:1971]	0.47
CRYAB	crystallin alpha B [Source:HGNC Symbol;Acc:HGNC:2389]	0.36
CTSO	cathepsin O [Source:HGNC Symbol;Acc:HGNC:2542]	-0.46
EHD2	EH domain containing 2 [Source:HGNC Symbol;Acc:HGNC:3243]	0.53
FLOT1	flotillin 1 [Source:HGNC Symbol;Acc:HGNC:3757]	0.42
GSN	gelsolin [Source:HGNC Symbol;Acc:HGNC:4620]	0.37

HIP1	huntingtin interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:4913]	0.40
HYAL2	hyaluronoglucosaminidase 2 [Source:HGNC Symbol;Acc:HGNC:5321]	0.42
INPP5F	inositol polyphosphate-5-phosphatase F [Source:HGNC Symbol;Acc:HGNC:17054]	0.01
MCOLN1	mucolipin 1 [Source:HGNC Symbol;Acc:HGNC:13356]	0.43
MYO1C	myosin IC [Source:HGNC Symbol;Acc:HGNC:7597]	0.50
OSBPL3	oxysterol binding protein like 3 [Source:HGNC Symbol;Acc:HGNC:16370]	-0.30
PI4KA	phosphatidylinositol 4-kinase alpha [Source:HGNC Symbol;Acc:HGNC:8983]	0.51
PIP5K1C	phosphatidylinositol-4-phosphate 5-kinase type 1 gamma [Source:HGNC Symbol;Acc:HGNC:8996]	0.62
PITPNA	phosphatidylinositol transfer protein alpha [Source:HGNC Symbol;Acc:HGNC:9001]	0.40
POLG	DNA polymerase gamma, catalytic subunit [Source:HGNC Symbol;Acc:HGNC:9179]	0.37
RAB35	RAB35, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9774]	0.64
RAB3D	RAB3D, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9779]	0.08
RHOB	ras homolog family member B [Source:HGNC Symbol;Acc:HGNC:668]	0.43
RIN3	Ras and Rab interactor 3 [Source:HGNC Symbol;Acc:HGNC:18751]	0.46
SCYL1	SCY1 like pseudokinase 1 [Source:HGNC Symbol;Acc:HGNC:14372]	0.76
SNX15	sorting nexin 15 [Source:HGNC Symbol;Acc:HGNC:14978]	0.48
STIM2	stromal interaction molecule 2 [Source:HGNC Symbol;Acc:HGNC:19205]	-0.47
SYT11	synaptotagmin 11 [Source:HGNC Symbol;Acc:HGNC:19239]	0.17
TBC1D25	TBC1 domain family member 25 [Source:HGNC Symbol;Acc:HGNC:8092]	0.54
TRAPPC12	trafficking protein particle complex 12 [Source:HGNC Symbol;Acc:HGNC:24284]	0.37
UNC45A	unc-45 myosin chaperone A [Source:HGNC Symbol;Acc:HGNC:30594]	0.54
WFS1	wolframin ER transmembrane glycoprotein [Source:HGNC Symbol;Acc:HGNC:12762]	0.46
XYLT1	xylosyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:15516]	0.39
Cons.11		
ACBD3	acyl-CoA binding domain containing 3 [Source:HGNC Symbol;Acc:HGNC:15453]	0.75
AP4M1	adaptor related protein complex 4 mu 1 subunit [Source:HGNC Symbol;Acc:HGNC:574]	-0.49
ARAP3	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 3 [Source:HGNC Symbol;Acc:HGNC:24097]	-0.31
ASAP3	ArfGAP with SH3 domain, ankyrin repeat and PH domain 3 [Source:HGNC Symbol;Acc:HGNC:14987]	-0.20
ATP6V0A2	ATPase H+ transporting V0 subunit a2 [Source:HGNC Symbol;Acc:HGNC:18481]	0.54
C1GALT1	core 1 synthase, glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:24337]	0.40

COG4	component of oligomeric golgi complex 4 [Source:HGNC Symbol;Acc:HGNC:18620]	-0.55
DGKZ	diacylglycerol kinase zeta [Source:HGNC Symbol;Acc:HGNC:2857]	-0.46
GALT	galactose-1-phosphate uridylyltransferase [Source:HGNC Symbol;Acc:HGNC:4135]	-0.71
GOLT1B	golgi transport 1B [Source:HGNC Symbol;Acc:HGNC:20175]	0.70
HSPA13	heat shock protein family A (Hsp70) member 13 [Source:HGNC Symbol;Acc:HGNC:11375]	0.75
IQSEC2	IQ motif and Sec7 domain 2 [Source:HGNC Symbol;Acc:HGNC:29059]	-0.47
KDSR	3-ketodihydrosphingosine reductase [Source:HGNC Symbol;Acc:HGNC:4021]	0.52
MAN1A1	mannosidase alpha class 1A member 1 [Source:HGNC Symbol;Acc:HGNC:6821]	0.39
MICAL1	microtubule associated monooxygenase, calponin and LIM domain containing 1 [Source:HGNC Symbol;Acc:HGNC:20619]	-0.22
MKS1	Meckel syndrome, type 1 [Source:HGNC Symbol;Acc:HGNC:7121]	-0.52
OCRL	OCRL, inositol polyphosphate-5-phosphatase [Source:HGNC Symbol;Acc:HGNC:8108]	0.37
PACS2	phosphofurin acidic cluster sorting protein 2 [Source:HGNC Symbol;Acc:HGNC:23794]	-0.52
PEX6	peroxisomal biogenesis factor 6 [Source:HGNC Symbol;Acc:HGNC:8859]	-0.43
PIGH	phosphatidylinositol glycan anchor biosynthesis class H [Source:HGNC Symbol;Acc:HGNC:8964]	0.33
PPAP2B	#N/A	0.40
RABGGTB	Rab geranylgeranyltransferase beta subunit [Source:HGNC Symbol;Acc:HGNC:9796]	0.48
RCN2	reticulocalbin 2 [Source:HGNC Symbol;Acc:HGNC:9935]	0.36
SEC63	SEC63 homolog, protein translocation regulator [Source:HGNC Symbol;Acc:HGNC:21082]	0.78
SEH1L	SEH1 like nucleoporin [Source:HGNC Symbol;Acc:HGNC:30379]	0.59
SLC27A1	solute carrier family 27 member 1 [Source:HGNC Symbol;Acc:HGNC:10995]	-0.70
SRP72	signal recognition particle 72 [Source:HGNC Symbol;Acc:HGNC:11303]	0.82
TBC1D13	TBC1 domain family member 13 [Source:HGNC Symbol;Acc:HGNC:25571]	-0.54
TRAM1	translocation associated membrane protein 1 [Source:HGNC Symbol;Acc:HGNC:20568]	0.71
TSC2	TSC complex subunit 2 [Source:HGNC Symbol;Acc:HGNC:12363]	-0.63
UGGT2	UDP-glucose glycoprotein glucosyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:15664]	0.56
VPS52	VPS52, GARP complex subunit [Source:HGNC Symbol;Acc:HGNC:10518]	-0.50
Cons.12*		
ACAP3	ArfGAP with coiled-coil, ankyrin repeat and PH domains 3 [Source:HGNC Symbol;Acc:HGNC:16754]	0.75
AGAP3	ArfGAP with GTPase domain, ankyrin repeat and PH domain 3 [Source:HGNC Symbol;Acc:HGNC:16923]	0.67
AP1G2	adaptor related protein complex 1 gamma 2 subunit [Source:HGNC	0.78

	Symbol;Acc:HGNC:556]	
ATG16L2	autophagy related 16 like 2 [Source:HGNC Symbol;Acc:HGNC:25464]	0.74
ATG4B	autophagy related 4B cysteine peptidase [Source:HGNC Symbol;Acc:HGNC:20790]	0.71
GGA1	golgi associated, gamma adaptin ear containing, ARF binding protein 1 [Source:HGNC Symbol;Acc:HGNC:17842]	0.84
GGA2	golgi associated, gamma adaptin ear containing, ARF binding protein 2 [Source:HGNC Symbol;Acc:HGNC:16064]	0.45
GGA3	golgi associated, gamma adaptin ear containing, ARF binding protein 3 [Source:HGNC Symbol;Acc:HGNC:17079]	0.74
IMPAD1	inositol monophosphatase domain containing 1 [Source:HGNC Symbol;Acc:HGNC:26019]	-0.68
INPP5E	inositol polyphosphate-5-phosphatase E [Source:HGNC Symbol;Acc:HGNC:21474]	0.63
LPCAT4	lysophosphatidylcholine acyltransferase 4 [Source:HGNC Symbol;Acc:HGNC:30059]	0.61
NBEAL2	neurobeachin like 2 [Source:HGNC Symbol;Acc:HGNC:31928]	0.60
NEK8	NIMA related kinase 8 [Source:HGNC Symbol;Acc:HGNC:13387]	0.46
OSBPL2	oxysterol binding protein like 2 [Source:HGNC Symbol;Acc:HGNC:15761]	0.49
OSBPL7	oxysterol binding protein like 7 [Source:HGNC Symbol;Acc:HGNC:16387]	0.74
PIK3R1	phosphoinositide-3-kinase regulatory subunit 1 [Source:HGNC Symbol;Acc:HGNC:8979]	-0.44
PRKD2	protein kinase D2 [Source:HGNC Symbol;Acc:HGNC:17293]	0.58
RAB24	RAB24, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9765]	0.73
RABEP2	rabaptin, RAB GTPase binding effector protein 2 [Source:HGNC Symbol;Acc:HGNC:24817]	0.52
RABGGTA	Rab geranylgeranyltransferase alpha subunit [Source:HGNC Symbol;Acc:HGNC:9795]	0.61
RUSC1	RUN and SH3 domain containing 1 [Source:HGNC Symbol;Acc:HGNC:17153]	0.54
STX10	syntaxin 10 [Source:HGNC Symbol;Acc:HGNC:11428]	0.74
STX4	syntaxin 4 [Source:HGNC Symbol;Acc:HGNC:11439]	0.58
TAZ	tafazzin [Source:HGNC Symbol;Acc:HGNC:11577]	0.78
ULK3	unc-51 like kinase 3 [Source:HGNC Symbol;Acc:HGNC:19703]	0.66
WHAMM	WAS protein homolog associated with actin, golgi membranes and microtubules [Source:HGNC Symbol;Acc:HGNC:30493]	0.36
ZFYVE27	zinc finger FYVE-type containing 27 [Source:HGNC Symbol;Acc:HGNC:26559]	0.62
Cons.13		
ACOX1	acyl-CoA oxidase 1 [Source:HGNC Symbol;Acc:HGNC:119]	0.63
AHI1	Abelson helper integration site 1 [Source:HGNC Symbol;Acc:HGNC:21575]	0.18
AKAP8	A-kinase anchoring protein 8 [Source:HGNC Symbol;Acc:HGNC:378]	0.44
AP1G1	adaptor related protein complex 1 gamma 1 subunit [Source:HGNC Symbol;Acc:HGNC:555]	0.81
ATP7B	ATPase copper transporting beta [Source:HGNC Symbol;Acc:HGNC:870]	-0.28

CLINT1	clathrin interactor 1 [Source:HGNC Symbol;Acc:HGNC:23186]	0.67
COG2	component of oligomeric golgi complex 2 [Source:HGNC Symbol;Acc:HGNC:6546]	0.42
DENND5A	DENN domain containing 5A [Source:HGNC Symbol;Acc:HGNC:19344]	-0.14
DYNLL1	dynein light chain LC8-type 1 [Source:HGNC Symbol;Acc:HGNC:15476]	-0.27
FGD4	FYVE, RhoGEF and PH domain containing 4 [Source:HGNC Symbol;Acc:HGNC:19125]	0.27
GAPVD1	GTPase activating protein and VPS9 domains 1 [Source:HGNC Symbol;Acc:HGNC:23375]	0.72
IFT46	intraflagellar transport 46 [Source:HGNC Symbol;Acc:HGNC:26146]	-0.41
KIF3C	kinesin family member 3C [Source:HGNC Symbol;Acc:HGNC:6321]	-0.52
LFNG	LFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase [Source:HGNC Symbol;Acc:HGNC:6560]	-0.29
LMAN2L	lectin, mannose binding 2 like [Source:HGNC Symbol;Acc:HGNC:19263]	-0.32
NECAP2	NECAP endocytosis associated 2 [Source:HGNC Symbol;Acc:HGNC:25528]	-0.28
NPHP1	nephrocystin 1 [Source:HGNC Symbol;Acc:HGNC:7905]	-0.28
OSBP	oxysterol binding protein [Source:HGNC Symbol;Acc:HGNC:8503]	0.67
PIP5K1A	phosphatidylinositol-4-phosphate 5-kinase type 1 alpha [Source:HGNC Symbol;Acc:HGNC:8994]	0.41
RAB11FIP4	RAB11 family interacting protein 4 [Source:HGNC Symbol;Acc:HGNC:30267]	0.11
SAR1B	secretion associated Ras related GTPase 1B [Source:HGNC Symbol;Acc:HGNC:10535]	0.66
SEC16A	SEC16 homolog A, endoplasmic reticulum export factor [Source:HGNC Symbol;Acc:HGNC:29006]	0.59
SRP68	signal recognition particle 68 [Source:HGNC Symbol;Acc:HGNC:11302]	0.58
STX3	syntaxin 3 [Source:HGNC Symbol;Acc:HGNC:11438]	0.31
SULF1	sulfatase 1 [Source:HGNC Symbol;Acc:HGNC:20391]	-0.30
TRAPPC10	trafficking protein particle complex 10 [Source:HGNC Symbol;Acc:HGNC:11868]	0.54
VPS37B	VPS37B, ESCRT-I subunit [Source:HGNC Symbol;Acc:HGNC:25754]	0.35
Cons.14*		
ATL3	atlastin GTPase 3 [Source:HGNC Symbol;Acc:HGNC:24526]	0.71
CAV1	caveolin 1 [Source:HGNC Symbol;Acc:HGNC:1527]	0.49
CAV2	caveolin 2 [Source:HGNC Symbol;Acc:HGNC:1528]	0.59
GOLPH3	golgi phosphoprotein 3 [Source:HGNC Symbol;Acc:HGNC:15452]	0.78
MAP1LC3B	microtubule associated protein 1 light chain 3 beta [Source:HGNC Symbol;Acc:HGNC:13352]	0.68
MTMR2	myotubularin related protein 2 [Source:HGNC Symbol;Acc:HGNC:7450]	0.61
PICALM	phosphatidylinositol binding clathrin assembly protein [Source:HGNC Symbol;Acc:HGNC:15514]	0.78
PIP4K2A	phosphatidylinositol-5-phosphate 4-kinase type 2 alpha [Source:HGNC Symbol;Acc:HGNC:8997]	0.45
PITPNB	phosphatidylinositol transfer protein beta [Source:HGNC	0.72

	Symbol;Acc:HGNC:9002]	
PLEKHA3	pleckstrin homology domain containing A3 [Source:HGNC Symbol;Acc:HGNC:14338]	0.74
RAB21	RAB21, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:18263]	0.76
RAB22A	RAB22A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9764]	0.60
RAB31	RAB31, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9771]	0.56
RAB5B	RAB5B, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9784]	0.50
RAB8B	RAB8B, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:30273]	0.67
RAC1	Rac family small GTPase 1 [Source:HGNC Symbol;Acc:HGNC:9801]	0.71
RPGR	retinitis pigmentosa GTPase regulator [Source:HGNC Symbol;Acc:HGNC:10295]	0.49
RRAGC	Ras related GTP binding C [Source:HGNC Symbol;Acc:HGNC:19902]	0.64
SEC62	SEC62 homolog, preprotein translocation factor [Source:HGNC Symbol;Acc:HGNC:11846]	0.61
SLC37A4	solute carrier family 37 member 4 [Source:HGNC Symbol;Acc:HGNC:4061]	-0.50
SPG20	#N/A	0.62
STX12	syntaxin 12 [Source:HGNC Symbol;Acc:HGNC:11430]	0.72
VAMP3	vesicle associated membrane protein 3 [Source:HGNC Symbol;Acc:HGNC:12644]	0.71
VPS37A	VPS37A, ESCRT-I subunit [Source:HGNC Symbol;Acc:HGNC:24928]	0.72
XYLT2	xylosyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:15517]	-0.64
Cons.15		
ARF3	ADP ribosylation factor 3 [Source:HGNC Symbol;Acc:HGNC:654]	0.74
ARF6	ADP ribosylation factor 6 [Source:HGNC Symbol;Acc:HGNC:659]	0.32
B4GALT5	beta-1,4-galactosyltransferase 5 [Source:HGNC Symbol;Acc:HGNC:928]	0.50
BACE1	beta-secretase 1 [Source:HGNC Symbol;Acc:HGNC:933]	0.51
CLEC2D	C-type lectin domain family 2 member D [Source:HGNC Symbol;Acc:HGNC:14351]	-0.62
DCTN1	dynactin subunit 1 [Source:HGNC Symbol;Acc:HGNC:2711]	0.78
DYNC1H1	dynein cytoplasmic 1 heavy chain 1 [Source:HGNC Symbol;Acc:HGNC:2961]	0.78
EHD1	EH domain containing 1 [Source:HGNC Symbol;Acc:HGNC:3242]	0.52
ERC1	ELKS/RAB6-interacting/CAST family member 1 [Source:HGNC Symbol;Acc:HGNC:17072]	0.70
EXTL3	exostosin like glycosyltransferase 3 [Source:HGNC Symbol;Acc:HGNC:3518]	0.72
GOLGA8B	golgin A8 family member B [Source:HGNC Symbol;Acc:HGNC:31973]	-0.55
INPP5A	inositol polyphosphate-5-phosphatase A [Source:HGNC Symbol;Acc:HGNC:6076]	0.54
KIF1C	kinesin family member 1C [Source:HGNC Symbol;Acc:HGNC:6317]	0.59
M6PR	mannose-6-phosphate receptor, cation dependent [Source:HGNC Symbol;Acc:HGNC:6752]	0.51
MBTPS1	membrane bound transcription factor peptidase, site 1	0.47

	[Source:HGNC Symbol;Acc:HGNC:15456]	
MICALL1	MICAL like 1 [Source:HGNC Symbol;Acc:HGNC:29804]	0.61
NFKB1	nuclear factor kappa B subunit 1 [Source:HGNC Symbol;Acc:HGNC:7794]	0.44
PEX26	peroxisomal biogenesis factor 26 [Source:HGNC Symbol;Acc:HGNC:22965]	0.66
RAB11FIP5	RAB11 family interacting protein 5 [Source:HGNC Symbol;Acc:HGNC:24845]	0.59
RAB8A	RAB8A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:7007]	0.63
SNPH	syntaphilin [Source:HGNC Symbol;Acc:HGNC:15931]	0.22
ST3GAL2	ST3 beta-galactoside alpha-2,3-sialyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:10863]	0.47
TGFBRAP1	transforming growth factor beta receptor associated protein 1 [Source:HGNC Symbol;Acc:HGNC:16836]	0.68
Cons.16*		
APPL2	adaptor protein, phosphotyrosine interacting with PH domain and leucine zipper 2 [Source:HGNC Symbol;Acc:HGNC:18242]	0.51
ARFGEF1	ADP ribosylation factor guanine nucleotide exchange factor 1 [Source:HGNC Symbol;Acc:HGNC:15772]	0.82
CPNE3	copine 3 [Source:HGNC Symbol;Acc:HGNC:2316]	0.82
EGFR	epidermal growth factor receptor [Source:HGNC Symbol;Acc:HGNC:3236]	0.40
EXOC4	exocyst complex component 4 [Source:HGNC Symbol;Acc:HGNC:30389]	0.43
EXOC6	exocyst complex component 6 [Source:HGNC Symbol;Acc:HGNC:23196]	0.47
GOSR1	golgi SNAP receptor complex member 1 [Source:HGNC Symbol;Acc:HGNC:4430]	0.58
LPCAT1	lysophosphatidylcholine acyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:25718]	-0.45
PIK3CB	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit beta [Source:HGNC Symbol;Acc:HGNC:8976]	0.70
PSEN1	presenilin 1 [Source:HGNC Symbol;Acc:HGNC:9508]	0.71
RAB3IL1	RAB3A interacting protein like 1 [Source:HGNC Symbol;Acc:HGNC:9780]	-0.53
RABEP1	rabaptin, RAB GTPase binding effector protein 1 [Source:HGNC Symbol;Acc:HGNC:17677]	0.43
SCYL2	SCY1 like pseudokinase 2 [Source:HGNC Symbol;Acc:HGNC:19286]	0.69
SCYL3	SCY1 like pseudokinase 3 [Source:HGNC Symbol;Acc:HGNC:19285]	0.52
SEC22C	SEC22 homolog C, vesicle trafficking protein [Source:HGNC Symbol;Acc:HGNC:16828]	0.50
SPAST	spastin [Source:HGNC Symbol;Acc:HGNC:11233]	0.63
SPTLC1	serine palmitoyltransferase long chain base subunit 1 [Source:HGNC Symbol;Acc:HGNC:11277]	0.75
STX1A	syntaxin 1A [Source:HGNC Symbol;Acc:HGNC:11433]	-0.61
STXBP5	syntaxin binding protein 5 [Source:HGNC Symbol;Acc:HGNC:19665]	0.57
TRAPPC11	trafficking protein particle complex 11 [Source:HGNC Symbol;Acc:HGNC:25751]	0.71
TRAPPC8	trafficking protein particle complex 8 [Source:HGNC	0.83

	Symbol;Acc:HGNC:29169]	
VPS4B	vacuolar protein sorting 4 homolog B [Source:HGNC Symbol;Acc:HGNC:10895]	0.81
Cons.17		
ACAP2	ArfGAP with coiled-coil, ankyrin repeat and PH domains 2 [Source:HGNC Symbol;Acc:HGNC:16469]	0.77
AGPS	alkylglycerone phosphate synthase [Source:HGNC Symbol;Acc:HGNC:327]	0.64
AP4E1	adaptor related protein complex 4 epsilon 1 subunit [Source:HGNC Symbol;Acc:HGNC:573]	0.64
ARFIP1	ADP ribosylation factor interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:21496]	0.68
ATP7A	ATPase copper transporting alpha [Source:HGNC Symbol;Acc:HGNC:869]	0.67
COG6	component of oligomeric golgi complex 6 [Source:HGNC Symbol;Acc:HGNC:18621]	0.62
COL4A3BP	collagen type IV alpha 3 binding protein [Source:HGNC Symbol;Acc:HGNC:2205]	0.67
HPS3	HPS3, biogenesis of lysosomal organelles complex 2 subunit 1 [Source:HGNC Symbol;Acc:HGNC:15597]	0.44
MTM1	myotubularin 1 [Source:HGNC Symbol;Acc:HGNC:7448]	0.68
OSBPL11	oxysterol binding protein like 11 [Source:HGNC Symbol;Acc:HGNC:16397]	0.58
OSBPL8	oxysterol binding protein like 8 [Source:HGNC Symbol;Acc:HGNC:16396]	0.69
PIK3C3	phosphatidylinositol 3-kinase catalytic subunit type 3 [Source:HGNC Symbol;Acc:HGNC:8974]	0.69
PIK3R4	phosphoinositide-3-kinase regulatory subunit 4 [Source:HGNC Symbol;Acc:HGNC:8982]	0.62
PRKAA1	protein kinase AMP-activated catalytic subunit alpha 1 [Source:HGNC Symbol;Acc:HGNC:9376]	0.65
SACM1L	SAC1 like phosphatidylinositide phosphatase [Source:HGNC Symbol;Acc:HGNC:17059]	0.72
SNX13	sorting nexin 13 [Source:HGNC Symbol;Acc:HGNC:21335]	0.76
STX7	syntaxin 7 [Source:HGNC Symbol;Acc:HGNC:11442]	0.55
STXBP3	syntaxin binding protein 3 [Source:HGNC Symbol;Acc:HGNC:11446]	0.75
TRIP11	thyroid hormone receptor interactor 11 [Source:HGNC Symbol;Acc:HGNC:12305]	0.68
VCPIP1	valosin containing protein interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:30897]	0.71
WDR44	WD repeat domain 44 [Source:HGNC Symbol;Acc:HGNC:30512]	0.74
YIPF4	Yip1 domain family member 4 [Source:HGNC Symbol;Acc:HGNC:28145]	0.68
Cons.18*		
ABCB7	ATP binding cassette subfamily B member 7 [Source:HGNC Symbol;Acc:HGNC:48]	0.55
ASAP1	ArfGAP with SH3 domain, ankyrin repeat and PH domain 1 [Source:HGNC Symbol;Acc:HGNC:2720]	0.45
ATP9A	ATPase phospholipid transporting 9A (putative) [Source:HGNC	0.52

	Symbol;Acc:HGNC:13540]	
BCS1L	BCS1 homolog, ubiquinol-cytochrome c reductase complex chaperone [Source:HGNC Symbol;Acc:HGNC:1020]	-0.74
CDKN1B	cyclin dependent kinase inhibitor 1B [Source:HGNC Symbol;Acc:HGNC:1785]	0.45
CDS2	CDP-diacylglycerol synthase 2 [Source:HGNC Symbol;Acc:HGNC:1801]	0.61
EPS15	epidermal growth factor receptor pathway substrate 15 [Source:HGNC Symbol;Acc:HGNC:3419]	0.72
GABARAPL1	GABA type A receptor associated protein like 1 [Source:HGNC Symbol;Acc:HGNC:4068]	0.59
GCC2	GRIP and coiled-coil domain containing 2 [Source:HGNC Symbol;Acc:HGNC:23218]	0.70
KIF1B	kinesin family member 1B [Source:HGNC Symbol;Acc:HGNC:16636]	0.62
OPTN	optineurin [Source:HGNC Symbol;Acc:HGNC:17142]	0.58
OSBPL9	oxysterol binding protein like 9 [Source:HGNC Symbol;Acc:HGNC:16386]	0.64
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha [Source:HGNC Symbol;Acc:HGNC:8975]	0.74
POMT2	protein O-mannosyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:19743]	-0.36
RAB12	RAB12, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:31332]	0.67
RB1CC1	RB1 inducible coiled-coil 1 [Source:HGNC Symbol;Acc:HGNC:15574]	0.83
SBF2	SET binding factor 2 [Source:HGNC Symbol;Acc:HGNC:2135]	0.72
SEC24B	SEC24 homolog B, COPII coat complex component [Source:HGNC Symbol;Acc:HGNC:10704]	0.77
SNX9	sorting nexin 9 [Source:HGNC Symbol;Acc:HGNC:14973]	0.50
TBC1D15	TBC1 domain family member 15 [Source:HGNC Symbol;Acc:HGNC:25694]	0.76
Cons.19		
AP1S2	adaptor related protein complex 1 sigma 2 subunit [Source:HGNC Symbol;Acc:HGNC:560]	0.53
ARL6IP5	ADP ribosylation factor like GTPase 6 interacting protein 5 [Source:HGNC Symbol;Acc:HGNC:16937]	0.69
ATP6AP2	ATPase H+ transporting accessory protein 2 [Source:HGNC Symbol;Acc:HGNC:18305]	0.70
CRTAP	cartilage associated protein [Source:HGNC Symbol;Acc:HGNC:2379]	0.53
CTSK	cathepsin K [Source:HGNC Symbol;Acc:HGNC:2536]	0.37
EXT2	exostosin glycosyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:3513]	0.56
GM2A	GM2 ganglioside activator [Source:HGNC Symbol;Acc:HGNC:4367]	0.55
GNS	glucosamine (N-acetyl)-6-sulfatase [Source:HGNC Symbol;Acc:HGNC:4422]	0.82
HEXB	hexosaminidase subunit beta [Source:HGNC Symbol;Acc:HGNC:4879]	0.74
ITSN1	intersectin 1 [Source:HGNC Symbol;Acc:HGNC:6183]	0.49
LAPTM4A	lysosomal protein transmembrane 4 alpha [Source:HGNC Symbol;Acc:HGNC:6924]	0.72
LIPA	lipase A, lysosomal acid type [Source:HGNC Symbol;Acc:HGNC:6617]	0.72

LPCAT2	lysophosphatidylcholine acyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:26032]	0.53
MANBA	mannosidase beta [Source:HGNC Symbol;Acc:HGNC:6831]	0.55
NPC2	NPC intracellular cholesterol transporter 2 [Source:HGNC Symbol;Acc:HGNC:14537]	0.54
PHYH	phytanoyl-CoA 2-hydroxylase [Source:HGNC Symbol;Acc:HGNC:8940]	0.42
PPT1	palmitoyl-protein thioesterase 1 [Source:HGNC Symbol;Acc:HGNC:9325]	0.77
SDCBP	syndecan binding protein [Source:HGNC Symbol;Acc:HGNC:10662]	0.58
TGOLN2	trans-golgi network protein 2 [Source:HGNC Symbol;Acc:HGNC:15450]	0.71
TPP1	tripeptidyl peptidase 1 [Source:HGNC Symbol;Acc:HGNC:2073]	0.72
Cons.20*		
ARL2	ADP ribosylation factor like GTPase 2 [Source:HGNC Symbol;Acc:HGNC:693]	0.77
ARL6IP4	ADP ribosylation factor like GTPase 6 interacting protein 4 [Source:HGNC Symbol;Acc:HGNC:18076]	0.85
B3GAT3	beta-1,3-glucuronyltransferase 3 [Source:HGNC Symbol;Acc:HGNC:923]	0.65
B4GALT7	beta-1,4-galactosyltransferase 7 [Source:HGNC Symbol;Acc:HGNC:930]	0.63
CHMP4A	charged multivesicular body protein 4A [Source:HGNC Symbol;Acc:HGNC:20274]	0.57
DPM3	dolichyl-phosphate mannosyltransferase subunit 3 [Source:HGNC Symbol;Acc:HGNC:3007]	0.76
HRAS	HRas proto-oncogene, GTPase [Source:HGNC Symbol;Acc:HGNC:5173]	0.69
LAMTOR4	late endosomal/lysosomal adaptor, MAPK and MTOR activator 4 [Source:HGNC Symbol;Acc:HGNC:33772]	0.83
NDUFS7	NADH:ubiquinone oxidoreductase core subunit S7 [Source:HGNC Symbol;Acc:HGNC:7714]	0.83
NDUFS8	NADH:ubiquinone oxidoreductase core subunit S8 [Source:HGNC Symbol;Acc:HGNC:7715]	0.72
NDUFV1	NADH:ubiquinone oxidoreductase core subunit V1 [Source:HGNC Symbol;Acc:HGNC:7716]	0.69
RAB4B	RAB4B, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9782]	0.65
RABAC1	Rab acceptor 1 [Source:HGNC Symbol;Acc:HGNC:9794]	0.60
SLC27A5	solute carrier family 27 member 5 [Source:HGNC Symbol;Acc:HGNC:10999]	0.64
TRAPPC2L	trafficking protein particle complex 2 like [Source:HGNC Symbol;Acc:HGNC:30887]	0.58
USE1	unconventional SNARE in the ER 1 [Source:HGNC Symbol;Acc:HGNC:30882]	0.64
VAMP5	vesicle associated membrane protein 5 [Source:HGNC Symbol;Acc:HGNC:12646]	0.48
VPS16	VPS16, CORVET/HOPS core subunit [Source:HGNC Symbol;Acc:HGNC:14584]	0.53
VPS28	VPS28, ESCRT-I subunit [Source:HGNC Symbol;Acc:HGNC:18178]	0.70

Cons.21		
AP2A2	adaptor related protein complex 2 alpha 2 subunit [Source:HGNC Symbol;Acc:HGNC:562]	-0.36
ARAP1	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1 [Source:HGNC Symbol;Acc:HGNC:16925]	-0.58
ATP6AP1L	ATPase H+ transporting accessory protein 1 like [Source:HGNC Symbol;Acc:HGNC:28091]	-0.38
CDK2	cyclin dependent kinase 2 [Source:HGNC Symbol;Acc:HGNC:1771]	0.07
DNM1L	dynamin 1 like [Source:HGNC Symbol;Acc:HGNC:2973]	0.69
GFM1	G elongation factor mitochondrial 1 [Source:HGNC Symbol;Acc:HGNC:13780]	0.86
LRPPRC	leucine rich pentatricopeptide repeat containing [Source:HGNC Symbol;Acc:HGNC:15714]	0.75
MAN2B2	mannosidase alpha class 2B member 2 [Source:HGNC Symbol;Acc:HGNC:29623]	-0.25
MTOR	mechanistic target of rapamycin kinase [Source:HGNC Symbol;Acc:HGNC:3942]	0.57
NDUFS1	NADH:ubiquinone oxidoreductase core subunit S1 [Source:HGNC Symbol;Acc:HGNC:7707]	0.82
PDHA1	pyruvate dehydrogenase E1 alpha 1 subunit [Source:HGNC Symbol;Acc:HGNC:8806]	0.63
PIP4K2B	phosphatidylinositol-5-phosphate 4-kinase type 2 beta [Source:HGNC Symbol;Acc:HGNC:8998]	-0.40
SCO1	SCO1, cytochrome c oxidase assembly protein [Source:HGNC Symbol;Acc:HGNC:10603]	0.63
SEC23IP	SEC23 interacting protein [Source:HGNC Symbol;Acc:HGNC:17018]	0.78
SNX1	sorting nexin 1 [Source:HGNC Symbol;Acc:HGNC:11172]	0.30
SRPR	#N/A	0.66
TRAPPC13	trafficking protein particle complex 13 [Source:HGNC Symbol;Acc:HGNC:25828]	0.67
TRAPPC2	trafficking protein particle complex 2 [Source:HGNC Symbol;Acc:HGNC:23068]	-0.32
UNC13D	unc-13 homolog D [Source:HGNC Symbol;Acc:HGNC:23147]	-0.56
Cons.22*		
AGAP6	ArfGAP with GTPase domain, ankyrin repeat and PH domain 6 [Source:HGNC Symbol;Acc:HGNC:23466]	0.61
AP5Z1	adaptor related protein complex 5 zeta 1 subunit [Source:HGNC Symbol;Acc:HGNC:22197]	0.70
ARFGAP1	ADP ribosylation factor GTPase activating protein 1 [Source:HGNC Symbol;Acc:HGNC:15852]	0.81
CLCN7	chloride voltage-gated channel 7 [Source:HGNC Symbol;Acc:HGNC:2025]	0.68
FIG4	FIG4 phosphoinositide 5-phosphatase [Source:HGNC Symbol;Acc:HGNC:16873]	-0.40
FYCO1	FYVE and coiled-coil domain containing 1 [Source:HGNC Symbol;Acc:HGNC:14673]	-0.45
IDUA	iduronidase, alpha-L- [Source:HGNC Symbol;Acc:HGNC:5391]	0.75
LEPRE1	#N/A	0.47
MAN1B1	mannosidase alpha class 1B member 1 [Source:HGNC	0.60

	Symbol;Acc:HGNC:6823]	
MOGS	mannosyl-oligosaccharide glucosidase [Source:HGNC Symbol;Acc:HGNC:24862]	0.73
OSBPL1A	oxysterol binding protein like 1A [Source:HGNC Symbol;Acc:HGNC:16398]	-0.65
POFUT2	protein O-fucosyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:14683]	0.71
PRKCA	protein kinase C alpha [Source:HGNC Symbol;Acc:HGNC:9393]	-0.39
SGSH	N-sulfoglucosamine sulfohydrolase [Source:HGNC Symbol;Acc:HGNC:10818]	0.44
SLC35C2	solute carrier family 35 member C2 [Source:HGNC Symbol;Acc:HGNC:17117]	0.70
SMPD4	sphingomyelin phosphodiesterase 4 [Source:HGNC Symbol;Acc:HGNC:32949]	0.54
SPG7	SPG7, paraplegin matrix AAA peptidase subunit [Source:HGNC Symbol;Acc:HGNC:11237]	0.65
TBC1D5	TBC1 domain family member 5 [Source:HGNC Symbol;Acc:HGNC:19166]	-0.53
ZFYVE9	zinc finger FYVE-type containing 9 [Source:HGNC Symbol;Acc:HGNC:6775]	-0.72
Cons.23*		
BBS2	Bardet-Biedl syndrome 2 [Source:HGNC Symbol;Acc:HGNC:967]	-0.36
CALR	calreticulin [Source:HGNC Symbol;Acc:HGNC:1455]	0.90
COPG1	coatomer protein complex subunit gamma 1 [Source:HGNC Symbol;Acc:HGNC:2236]	0.83
DNAJB11	DnaJ heat shock protein family (Hsp40) member B11 [Source:HGNC Symbol;Acc:HGNC:14889]	0.90
DPAGT1	dolichyl-phosphate N-acetylglucosaminephosphotransferase 1 [Source:HGNC Symbol;Acc:HGNC:2995]	0.71
GANAB	glucosidase II alpha subunit [Source:HGNC Symbol;Acc:HGNC:4138]	0.74
GBF1	golgi brefeldin A resistant guanine nucleotide exchange factor 1 [Source:HGNC Symbol;Acc:HGNC:4181]	0.41
HSP90B1	heat shock protein 90 beta family member 1 [Source:HGNC Symbol;Acc:HGNC:12028]	0.87
HSPA5	heat shock protein family A (Hsp70) member 5 [Source:HGNC Symbol;Acc:HGNC:5238]	0.84
ORMDL2	ORMDL sphingolipid biosynthesis regulator 2 [Source:HGNC Symbol;Acc:HGNC:16037]	0.68
PDIA3	protein disulfide isomerase family A member 3 [Source:HGNC Symbol;Acc:HGNC:4606]	0.88
PDIA4	protein disulfide isomerase family A member 4 [Source:HGNC Symbol;Acc:HGNC:30167]	0.90
PDIA6	protein disulfide isomerase family A member 6 [Source:HGNC Symbol;Acc:HGNC:30168]	0.93
SEC23B	Sec23 homolog B, coat complex II component [Source:HGNC Symbol;Acc:HGNC:10702]	0.77
SRPRB	SRP receptor beta subunit [Source:HGNC Symbol;Acc:HGNC:24085]	0.85
TIMM8A	translocase of inner mitochondrial membrane 8A [Source:HGNC Symbol;Acc:HGNC:11817]	0.40
UGGT1	UDP-glucose glycoprotein glucosyltransferase 1 [Source:HGNC	0.80

	Symbol;Acc:HGNC:15663]	
XBP1	X-box binding protein 1 [Source:HGNC Symbol;Acc:HGNC:12801]	0.70
Cons.24*		
AGAP1	ArfGAP with GTPase domain, ankyrin repeat and PH domain 1 [Source:HGNC Symbol;Acc:HGNC:16922]	0.32
AP4B1	adaptor related protein complex 4 beta 1 subunit [Source:HGNC Symbol;Acc:HGNC:572]	-0.46
APPL1	adaptor protein, phosphotyrosine interacting with PH domain and leucine zipper 1 [Source:HGNC Symbol;Acc:HGNC:24035]	0.82
ARL16	ADP ribosylation factor like GTPase 16 [Source:HGNC Symbol;Acc:HGNC:27902]	-0.48
B3GALNT2	beta-1,3-N-acetylgalactosaminyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:28596]	0.59
COG3	component of oligomeric golgi complex 3 [Source:HGNC Symbol;Acc:HGNC:18619]	0.70
COG5	component of oligomeric golgi complex 5 [Source:HGNC Symbol;Acc:HGNC:14857]	0.62
EEA1	early endosome antigen 1 [Source:HGNC Symbol;Acc:HGNC:3185]	0.85
EXOC5	exocyst complex component 5 [Source:HGNC Symbol;Acc:HGNC:10696]	0.85
GOLGA4	golgin A4 [Source:HGNC Symbol;Acc:HGNC:4427]	0.65
HOOK3	hook microtubule tethering protein 3 [Source:HGNC Symbol;Acc:HGNC:23576]	0.69
KIF2A	kinesin family member 2A [Source:HGNC Symbol;Acc:HGNC:6318]	0.69
MACF1	microtubule-actin crosslinking factor 1 [Source:HGNC Symbol;Acc:HGNC:13664]	0.38
RAB3GAP1	RAB3 GTPase activating protein catalytic subunit 1 [Source:HGNC Symbol;Acc:HGNC:17063]	0.72
RAB5A	RAB5A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9783]	0.80
SCAPER	S-phase cyclin A associated protein in the ER [Source:HGNC Symbol;Acc:HGNC:13081]	0.57
SMAP1	small ArfGAP 1 [Source:HGNC Symbol;Acc:HGNC:19651]	0.68
SNX29	sorting nexin 29 [Source:HGNC Symbol;Acc:HGNC:30542]	0.35
Cons.25		
ATP6V0E1	ATPase H+ transporting V0 subunit e1 [Source:HGNC Symbol;Acc:HGNC:863]	0.61
ATP6V1D	ATPase H+ transporting V1 subunit D [Source:HGNC Symbol;Acc:HGNC:13527]	0.52
ATP6V1E1	ATPase H+ transporting V1 subunit E1 [Source:HGNC Symbol;Acc:HGNC:857]	0.84
ATP6V1H	ATPase H+ transporting V1 subunit H [Source:HGNC Symbol;Acc:HGNC:18303]	0.71
BECN1	beclin 1 [Source:HGNC Symbol;Acc:HGNC:1034]	0.73
CHMP3	charged multivesicular body protein 3 [Source:HGNC Symbol;Acc:HGNC:29865]	0.63
OFD1	OFD1, centriole and centriolar satellite protein [Source:HGNC Symbol;Acc:HGNC:2567]	-0.65
RAB5C	RAB5C, member RAS oncogene family [Source:HGNC	0.69

	Symbol;Acc:HGNC:9785]	
RAB7A	RAB7A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9788]	0.85
RRAGA	Ras related GTP binding A [Source:HGNC Symbol;Acc:HGNC:16963]	0.75
SEC31B	SEC31 homolog B, COPII coat complex component [Source:HGNC Symbol;Acc:HGNC:23197]	-0.66
SNAP29	synaptosome associated protein 29 [Source:HGNC Symbol;Acc:HGNC:11133]	0.69
SORT1	sortilin 1 [Source:HGNC Symbol;Acc:HGNC:11186]	0.59
STX16	syntaxin 16 [Source:HGNC Symbol;Acc:HGNC:11431]	-0.70
TBC1D1	TBC1 domain family member 1 [Source:HGNC Symbol;Acc:HGNC:11578]	0.38
Cons.26		
ABCD1	ATP binding cassette subfamily D member 1 [Source:HGNC Symbol;Acc:HGNC:61]	0.63
ACP2	acid phosphatase 2, lysosomal [Source:HGNC Symbol;Acc:HGNC:123]	0.67
ALG1	ALG1, chitobiosyldiphosphodolichol beta-mannosyltransferase [Source:HGNC Symbol;Acc:HGNC:18294]	0.73
ARFGAP2	ADP ribosylation factor GTPase activating protein 2 [Source:HGNC Symbol;Acc:HGNC:13504]	0.56
ARRDC3	arrestin domain containing 3 [Source:HGNC Symbol;Acc:HGNC:29263]	-0.49
COG8	component of oligomeric golgi complex 8 [Source:HGNC Symbol;Acc:HGNC:18623]	0.58
DOLK	dolichol kinase [Source:HGNC Symbol;Acc:HGNC:23406]	0.81
HPS6	HPS6, biogenesis of lysosomal organelles complex 2 subunit 3 [Source:HGNC Symbol;Acc:HGNC:18817]	0.83
PSEN2	presenilin 2 [Source:HGNC Symbol;Acc:HGNC:9509]	0.56
PSKH1	protein serine kinase H1 [Source:HGNC Symbol;Acc:HGNC:9529]	0.73
RPTOR	regulatory associated protein of MTOR complex 1 [Source:HGNC Symbol;Acc:HGNC:30287]	0.66
SMPD1	sphingomyelin phosphodiesterase 1 [Source:HGNC Symbol;Acc:HGNC:11120]	0.59
VAC14	Vac14, PIKFYVE complex component [Source:HGNC Symbol;Acc:HGNC:25507]	0.69
VPS18	VPS18, CORVET/HOPS core subunit [Source:HGNC Symbol;Acc:HGNC:15972]	0.75
Cons.27		
AKAP9	A-kinase anchoring protein 9 [Source:HGNC Symbol;Acc:HGNC:379]	0.81
ARFGEF2	ADP ribosylation factor guanine nucleotide exchange factor 2 [Source:HGNC Symbol;Acc:HGNC:15853]	0.67
ATG14	autophagy related 14 [Source:HGNC Symbol;Acc:HGNC:19962]	0.73
ATG2B	autophagy related 2B [Source:HGNC Symbol;Acc:HGNC:20187]	0.81
CLASP1	cytoplasmic linker associated protein 1 [Source:HGNC Symbol;Acc:HGNC:17088]	0.59
EPG5	ectopic P-granules autophagy protein 5 homolog [Source:HGNC Symbol;Acc:HGNC:29331]	0.65
FGD6	FYVE, RhoGEF and PH domain containing 6 [Source:HGNC	0.46

	Symbol;Acc:HGNC:21740]	
GOLGB1	golgin B1 [Source:HGNC Symbol;Acc:HGNC:4429]	0.73
PEX1	peroxisomal biogenesis factor 1 [Source:HGNC Symbol;Acc:HGNC:8850]	0.72
PIGG	phosphatidylinositol glycan anchor biosynthesis class G [Source:HGNC Symbol;Acc:HGNC:25985]	0.58
SYNRG	synergin gamma [Source:HGNC Symbol;Acc:HGNC:557]	0.61
TNKS	tankyrase [Source:HGNC Symbol;Acc:HGNC:11941]	0.79
VPS13A	vacuolar protein sorting 13 homolog A [Source:HGNC Symbol;Acc:HGNC:1908]	0.63
VPS39	VPS39, HOPS complex subunit [Source:HGNC Symbol;Acc:HGNC:20593]	0.55
Cons.28		
CLTA	clathrin light chain A [Source:HGNC Symbol;Acc:HGNC:2090]	0.66
CTSD	cathepsin D [Source:HGNC Symbol;Acc:HGNC:2529]	0.73
ERGIC3	ERGIC and golgi 3 [Source:HGNC Symbol;Acc:HGNC:15927]	0.75
ERP29	endoplasmic reticulum protein 29 [Source:HGNC Symbol;Acc:HGNC:13799]	0.76
GNPTG	N-acetylglucosamine-1-phosphate transferase gamma subunit [Source:HGNC Symbol;Acc:HGNC:23026]	0.66
GUSB	glucuronidase beta [Source:HGNC Symbol;Acc:HGNC:4696]	0.51
IQCB1	IQ motif containing B1 [Source:HGNC Symbol;Acc:HGNC:28949]	-0.36
KDELR1	KDEL endoplasmic reticulum protein retention receptor 1 [Source:HGNC Symbol;Acc:HGNC:6304]	0.70
LRP5	LDL receptor related protein 5 [Source:HGNC Symbol;Acc:HGNC:6697]	0.30
LRPAP1	LDL receptor related protein associated protein 1 [Source:HGNC Symbol;Acc:HGNC:6701]	0.62
NEU1	neuraminidase 1 [Source:HGNC Symbol;Acc:HGNC:7758]	0.72
SQSTM1	sequestosome 1 [Source:HGNC Symbol;Acc:HGNC:11280]	0.45
TBC1D2	TBC1 domain family member 2 [Source:HGNC Symbol;Acc:HGNC:18026]	0.29
YIPF3	Yip1 domain family member 3 [Source:HGNC Symbol;Acc:HGNC:21023]	0.82
Cons.29		
ARL13B	ADP ribosylation factor like GTPase 13B [Source:HGNC Symbol;Acc:HGNC:25419]	0.66
ATG12	autophagy related 12 [Source:HGNC Symbol;Acc:HGNC:588]	0.69
BBS7	Bardet-Biedl syndrome 7 [Source:HGNC Symbol;Acc:HGNC:18758]	0.60
CLN5	CLN5, intracellular trafficking protein [Source:HGNC Symbol;Acc:HGNC:2076]	0.38
LYST	lysosomal trafficking regulator [Source:HGNC Symbol;Acc:HGNC:1968]	0.64
MYO5A	myosin VA [Source:HGNC Symbol;Acc:HGNC:7602]	0.53
PIK3C2A	phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2 alpha [Source:HGNC Symbol;Acc:HGNC:8971]	0.82
PIKFYVE	phosphoinositide kinase, FYVE-type zinc finger containing [Source:HGNC Symbol;Acc:HGNC:23785]	0.76
RAB11FIP	RAB11 family interacting protein 2 [Source:HGNC	0.78

2	Symbol;Acc:HGNC:29152]	
SYNJ1	synaptojanin 1 [Source:HGNC Symbol;Acc:HGNC:11503]	0.58
TGFBR3	transforming growth factor beta receptor 3 [Source:HGNC Symbol;Acc:HGNC:11774]	0.56
VAMP4	vesicle associated membrane protein 4 [Source:HGNC Symbol;Acc:HGNC:12645]	0.74
VPS54	VPS54, GARP complex subunit [Source:HGNC Symbol;Acc:HGNC:18652]	0.68
Cons.30		
ATP6AP1	ATPase H+ transporting accessory protein 1 [Source:HGNC Symbol;Acc:HGNC:868]	0.74
CTSA	cathepsin A [Source:HGNC Symbol;Acc:HGNC:9251]	0.75
FUCA2	alpha-L-fucosidase 2 [Source:HGNC Symbol;Acc:HGNC:4008]	0.65
GBA	glucosylceramidase beta [Source:HGNC Symbol;Acc:HGNC:4177]	0.77
GLB1	galactosidase beta 1 [Source:HGNC Symbol;Acc:HGNC:4298]	0.85
HEXA	hexosaminidase subunit alpha [Source:HGNC Symbol;Acc:HGNC:4878]	0.61
LAMP1	lysosomal associated membrane protein 1 [Source:HGNC Symbol;Acc:HGNC:6499]	0.77
MAN2B1	mannosidase alpha class 2B member 1 [Source:HGNC Symbol;Acc:HGNC:6826]	0.65
NAGA	alpha-N-acetylgalactosaminidase [Source:HGNC Symbol;Acc:HGNC:7631]	0.67
NCSTN	nicastrin [Source:HGNC Symbol;Acc:HGNC:17091]	0.73
POFUT1	protein O-fucosyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:14988]	0.57
PSAP	prosaposin [Source:HGNC Symbol;Acc:HGNC:9498]	0.76
SUMF1	sulfatase modifying factor 1 [Source:HGNC Symbol;Acc:HGNC:20376]	0.60
Cons.31		
ACAP1	ArfGAP with coiled-coil, ankyrin repeat and PH domains 1 [Source:HGNC Symbol;Acc:HGNC:16467]	0.52
ARRB2	arrestin beta 2 [Source:HGNC Symbol;Acc:HGNC:712]	0.76
CD68	CD68 molecule [Source:HGNC Symbol;Acc:HGNC:1693]	0.75
CTSB	cathepsin B [Source:HGNC Symbol;Acc:HGNC:2527]	0.65
CTSC	cathepsin C [Source:HGNC Symbol;Acc:HGNC:2528]	0.53
CTSS	cathepsin S [Source:HGNC Symbol;Acc:HGNC:2545]	0.77
FGD2	FYVE, RhoGEF and PH domain containing 2 [Source:HGNC Symbol;Acc:HGNC:3664]	0.72
FGD3	FYVE, RhoGEF and PH domain containing 3 [Source:HGNC Symbol;Acc:HGNC:16027]	0.58
INPP5D	inositol polyphosphate-5-phosphatase D [Source:HGNC Symbol;Acc:HGNC:6079]	0.48
LAPTM5	lysosomal protein transmembrane 5 [Source:HGNC Symbol;Acc:HGNC:29612]	0.86
SERPINA1	serpin family A member 1 [Source:HGNC Symbol;Acc:HGNC:8941]	0.52
SLC11A1	solute carrier family 11 member 1 [Source:HGNC Symbol;Acc:HGNC:10907]	0.59

Cons.32		
EXOC1	exocyst complex component 1 [Source:HGNC Symbol;Acc:HGNC:30380]	0.71
FKTN	fukutin [Source:HGNC Symbol;Acc:HGNC:3622]	0.69
MYO6	myosin VI [Source:HGNC Symbol;Acc:HGNC:7605]	0.59
PEX3	peroxisomal biogenesis factor 3 [Source:HGNC Symbol;Acc:HGNC:8858]	0.64
PI4K2B	phosphatidylinositol 4-kinase type 2 beta [Source:HGNC Symbol;Acc:HGNC:18215]	0.64
RAB28	RAB28, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9768]	0.40
SEC22A	SEC22 homolog A, vesicle trafficking protein [Source:HGNC Symbol;Acc:HGNC:20260]	0.48
SLC35A3	solute carrier family 35 member A3 [Source:HGNC Symbol;Acc:HGNC:11023]	0.83
SNX14	sorting nexin 14 [Source:HGNC Symbol;Acc:HGNC:14977]	0.73
TRAPPC6B	trafficking protein particle complex 6B [Source:HGNC Symbol;Acc:HGNC:23066]	0.78
VPS36	vacuolar protein sorting 36 homolog [Source:HGNC Symbol;Acc:HGNC:20312]	0.72
Cons.33		
DHCR24	24-dehydrocholesterol reductase [Source:HGNC Symbol;Acc:HGNC:2859]	0.75
DHCR7	7-dehydrocholesterol reductase [Source:HGNC Symbol;Acc:HGNC:2860]	0.85
EBP	emopamil binding protein (sterol isomerase) [Source:HGNC Symbol;Acc:HGNC:3133]	0.71
HMGCR	3-hydroxy-3-methylglutaryl-CoA reductase [Source:HGNC Symbol;Acc:HGNC:5006]	0.80
INSIG1	insulin induced gene 1 [Source:HGNC Symbol;Acc:HGNC:6083]	0.78
LDLR	low density lipoprotein receptor [Source:HGNC Symbol;Acc:HGNC:6547]	0.69
PRKCD	protein kinase C delta [Source:HGNC Symbol;Acc:HGNC:9399]	0.27
SREBF2	sterol regulatory element binding transcription factor 2 [Source:HGNC Symbol;Acc:HGNC:11290]	0.75
UNC13B	unc-13 homolog B [Source:HGNC Symbol;Acc:HGNC:12566]	0.23
Cons.34		
ATP6V1F	ATPase H+ transporting V1 subunit F [Source:HGNC Symbol;Acc:HGNC:16832]	-0.67
CEP290	centrosomal protein 290 [Source:HGNC Symbol;Acc:HGNC:29021]	0.78
ERO1LB	#N/A	0.61
GOPC	golgi associated PDZ and coiled-coil motif containing [Source:HGNC Symbol;Acc:HGNC:17643]	0.81
KIF3A	kinesin family member 3A [Source:HGNC Symbol;Acc:HGNC:6319]	0.67
MON2	MON2 homolog, regulator of endosome-to-Golgi trafficking [Source:HGNC Symbol;Acc:HGNC:29177]	0.79
NPHP3	nephrocystin 3 [Source:HGNC Symbol;Acc:HGNC:7907]	0.75
RUFY2	RUN and FYVE domain containing 2 [Source:HGNC	0.79

	Symbol;Acc:HGNC:19761]	
ZFYVE16	zinc finger FYVE-type containing 16 [Source:HGNC Symbol;Acc:HGNC:20756]	0.72
Cons.35		
DGUOK	deoxyguanosine kinase [Source:HGNC Symbol;Acc:HGNC:2858]	0.66
DYNLT1	dynein light chain Tctex-type 1 [Source:HGNC Symbol;Acc:HGNC:11697]	0.49
LAMTOR5	late endosomal/lysosomal adaptor, MAPK and MTOR activator 5 [Source:HGNC Symbol;Acc:HGNC:17955]	0.84
NDUFAF2	NADH:ubiquinone oxidoreductase complex assembly factor 2 [Source:HGNC Symbol;Acc:HGNC:28086]	0.84
NDUFV2	NADH:ubiquinone oxidoreductase core subunit V2 [Source:HGNC Symbol;Acc:HGNC:7717]	0.71
PARK7	Parkinsonism associated deglycase [Source:HGNC Symbol;Acc:HGNC:16369]	0.82
RABEPK	Rab9 effector protein with kelch motifs [Source:HGNC Symbol;Acc:HGNC:16896]	0.59
UQCRB	ubiquinol-cytochrome c reductase binding protein [Source:HGNC Symbol;Acc:HGNC:12582]	0.54
Cons.36		
ARPC2	actin related protein 2/3 complex subunit 2 [Source:HGNC Symbol;Acc:HGNC:705]	0.83
ARPC3	actin related protein 2/3 complex subunit 3 [Source:HGNC Symbol;Acc:HGNC:706]	0.80
ARPC5	actin related protein 2/3 complex subunit 5 [Source:HGNC Symbol;Acc:HGNC:708]	0.81
DEGS1	delta 4-desaturase, sphingolipid 1 [Source:HGNC Symbol;Acc:HGNC:13709]	0.73
RAB32	RAB32, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9772]	0.63
RTN4	reticulon 4 [Source:HGNC Symbol;Acc:HGNC:14085]	0.64

*** Consensus modules significant for all four preservation statistics**

Supplementary Table 2

TF	TF.description	Score
Cons.2		
CENPB	centromere protein B [Source:HGNC Symbol;Acc:HGNC:1852]	2.25
CREB3L2	cAMP responsive element binding protein 3 like 2 [Source:HGNC Symbol;Acc:HGNC:23720]	2.23
ELK3	ELK3, ETS transcription factor [Source:HGNC Symbol;Acc:HGNC:3325]	1.64
ERF	ETS2 repressor factor [Source:HGNC Symbol;Acc:HGNC:3444]	1.47
ETS2	ETS proto-oncogene 2, transcription factor [Source:HGNC	1.86

	Symbol;Acc:HGNC:3489]	
KLF13	Kruppel like factor 13 [Source:HGNC Symbol;Acc:HGNC:13672]	1.76
PKNOX1	PBX/knotted 1 homeobox 1 [Source:HGNC Symbol;Acc:HGNC:9022]	2.72
TGIF1	TGFB induced factor homeobox 1 [Source:HGNC Symbol;Acc:HGNC:11776]	1.73
ZBTB18	zinc finger and BTB domain containing 18 [Source:HGNC Symbol;Acc:HGNC:13030]	2.00
ZNF691	zinc finger protein 691 [Source:HGNC Symbol;Acc:HGNC:28028]	1.63
Cons.3		
CEBPZ	CCAAT/enhancer binding protein zeta [Source:HGNC Symbol;Acc:HGNC:24218]	1.76
CREB3L2	cAMP responsive element binding protein 3 like 2 [Source:HGNC Symbol;Acc:HGNC:23720]	2.03
EPAS1	endothelial PAS domain protein 1 [Source:HGNC Symbol;Acc:HGNC:3374]	5.15
HIC2	HIC ZBTB transcriptional repressor 2 [Source:HGNC Symbol;Acc:HGNC:18595]	3.99
KLF13	Kruppel like factor 13 [Source:HGNC Symbol;Acc:HGNC:13672]	1.42
NFKB2	nuclear factor kappa B subunit 2 [Source:HGNC Symbol;Acc:HGNC:7795]	1.46
NR4A1	nuclear receptor subfamily 4 group A member 1 [Source:HGNC Symbol;Acc:HGNC:7980]	1.53
RBPJ	recombination signal binding protein for immunoglobulin kappa J region [Source:HGNC Symbol;Acc:HGNC:5724]	1.48
RELB	RELB proto-oncogene, NF-kB subunit [Source:HGNC Symbol;Acc:HGNC:9956]	1.51
ZNF589	zinc finger protein 589 [Source:HGNC Symbol;Acc:HGNC:16747]	1.97
Cons.6		
ARNTL	aryl hydrocarbon receptor nuclear translocator like [Source:HGNC Symbol;Acc:HGNC:701]	1.39
BRF1	BRF1, RNA polymerase III transcription initiation factor subunit [Source:HGNC Symbol;Acc:HGNC:11551]	1.41
CENPB	centromere protein B [Source:HGNC Symbol;Acc:HGNC:1852]	1.53
CREB3L2	cAMP responsive element binding protein 3 like 2 [Source:HGNC Symbol;Acc:HGNC:23720]	1.54
FOXK1	forkhead box K1 [Source:HGNC Symbol;Acc:HGNC:23480]	1.32
HIC2	HIC ZBTB transcriptional repressor 2 [Source:HGNC Symbol;Acc:HGNC:18595]	3.27
NR1H3	nuclear receptor subfamily 1 group H member 3 [Source:HGNC Symbol;Acc:HGNC:7966]	2.54
STRA13	#N/A	1.34
TERF1	telomeric repeat binding factor 1 [Source:HGNC Symbol;Acc:HGNC:11728]	1.56
TFAP4	transcription factor AP-4 [Source:HGNC Symbol;Acc:HGNC:11745]	1.58
Cons.8		
CEBPZ	CCAAT/enhancer binding protein zeta [Source:HGNC Symbol;Acc:HGNC:24218]	1.69
DR1	down-regulator of transcription 1 [Source:HGNC	1.76

	Symbol;Acc:HGNC:3017]	
E2F3	E2F transcription factor 3 [Source:HGNC Symbol;Acc:HGNC:3115]	1.71
HIC2	HIC ZBTB transcriptional repressor 2 [Source:HGNC Symbol;Acc:HGNC:18595]	1.68
POU6F1	POU class 6 homeobox 1 [Source:HGNC Symbol;Acc:HGNC:9224]	1.65
RBPJ	recombination signal binding protein for immunoglobulin kappa J region [Source:HGNC Symbol;Acc:HGNC:5724]	1.94
STAT1	signal transducer and activator of transcription 1 [Source:HGNC Symbol;Acc:HGNC:11362]	1.69
STAT5A	signal transducer and activator of transcription 5A [Source:HGNC Symbol;Acc:HGNC:11366]	2.78
STAT5B	signal transducer and activator of transcription 5B [Source:HGNC Symbol;Acc:HGNC:11367]	2.16
TERF1	telomeric repeat binding factor 1 [Source:HGNC Symbol;Acc:HGNC:11728]	2.19
Cons.9		
BRCA1	BRCA1, DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1100]	1.84
CDC5L	cell division cycle 5 like [Source:HGNC Symbol;Acc:HGNC:1743]	1.95
CENPB	centromere protein B [Source:HGNC Symbol;Acc:HGNC:1852]	2.06
E4F1	E4F transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:3121]	1.70
EPAS1	endothelial PAS domain protein 1 [Source:HGNC Symbol;Acc:HGNC:3374]	17.77
GMEB2	glucocorticoid modulatory element binding protein 2 [Source:HGNC Symbol;Acc:HGNC:4371]	2.60
NF1	neurofibromin 1 [Source:HGNC Symbol;Acc:HGNC:7765]	1.64
ZBTB4	zinc finger and BTB domain containing 4 [Source:HGNC Symbol;Acc:HGNC:23847]	3.71
ZNF589	zinc finger protein 589 [Source:HGNC Symbol;Acc:HGNC:16747]	4.56
ZNF628	zinc finger protein 628 [Source:HGNC Symbol;Acc:HGNC:28054]	7.55
Cons.12		
CEBPD	CCAAT/enhancer binding protein delta [Source:HGNC Symbol;Acc:HGNC:1835]	1.40
CEBPG	CCAAT/enhancer binding protein gamma [Source:HGNC Symbol;Acc:HGNC:1837]	1.23
EP300	E1A binding protein p300 [Source:HGNC Symbol;Acc:HGNC:3373]	2.10
KLF12	Kruppel like factor 12 [Source:HGNC Symbol;Acc:HGNC:6346]	1.25
MNT	MAX network transcriptional repressor [Source:HGNC Symbol;Acc:HGNC:7188]	1.33
NFIC	nuclear factor I C [Source:HGNC Symbol;Acc:HGNC:7786]	1.67
RFX7	regulatory factor X7 [Source:HGNC Symbol;Acc:HGNC:25777]	2.07
STAT3	signal transducer and activator of transcription 3 [Source:HGNC Symbol;Acc:HGNC:11364]	1.34
TP53	tumor protein p53 [Source:HGNC Symbol;Acc:HGNC:11998]	1.40
USF1	upstream transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:12593]	1.50
Cons.14		
ATF1	activating transcription factor 1 [Source:HGNC	1.30

	Symbol;Acc:HGNC:783]	
CUX1	cut like homeobox 1 [Source:HGNC Symbol;Acc:HGNC:2557]	1.31
HMBOX1	homeobox containing 1 [Source:HGNC Symbol;Acc:HGNC:26137]	3.27
RARG	retinoic acid receptor gamma [Source:HGNC Symbol;Acc:HGNC:9866]	1.37
TRIM28	tripartite motif containing 28 [Source:HGNC Symbol;Acc:HGNC:16384]	1.40
TTF1	transcription termination factor 1 [Source:HGNC Symbol;Acc:HGNC:12397]	1.60
USF1	upstream transcription factor 1 [Source:HGNC Symbol;Acc:HGNC:12593]	2.03
ZBTB33	zinc finger and BTB domain containing 33 [Source:HGNC Symbol;Acc:HGNC:16682]	1.48
ZBTB49	zinc finger and BTB domain containing 49 [Source:HGNC Symbol;Acc:HGNC:19883]	1.75
ZBTB6	zinc finger and BTB domain containing 6 [Source:HGNC Symbol;Acc:HGNC:16764]	1.87
Cons.16		
ARID5A	AT-rich interaction domain 5A [Source:HGNC Symbol;Acc:HGNC:17361]	1.52
ATF5	activating transcription factor 5 [Source:HGNC Symbol;Acc:HGNC:790]	1.57
BBX	BBX, HMG-box containing [Source:HGNC Symbol;Acc:HGNC:14422]	1.68
CEBPB	CCAAT/enhancer binding protein beta [Source:HGNC Symbol;Acc:HGNC:1834]	1.89
ETS1	ETS proto-oncogene 1, transcription factor [Source:HGNC Symbol;Acc:HGNC:3488]	1.67
HMGA1	high mobility group AT-hook 1 [Source:HGNC Symbol;Acc:HGNC:5010]	1.76
PKNOX1	PBX/knotted 1 homeobox 1 [Source:HGNC Symbol;Acc:HGNC:9022]	2.18
REST	RE1 silencing transcription factor [Source:HGNC Symbol;Acc:HGNC:9966]	1.58
TGIF2	TGFB induced factor homeobox 2 [Source:HGNC Symbol;Acc:HGNC:15764]	1.73
ZNF350	zinc finger protein 350 [Source:HGNC Symbol;Acc:HGNC:16656]	1.73
Cons.18		
CTCF	CCCTC-binding factor [Source:HGNC Symbol;Acc:HGNC:13723]	1.54
GABPB1	GA binding protein transcription factor beta subunit 1 [Source:HGNC Symbol;Acc:HGNC:4074]	1.38
MAFF	MAF bZIP transcription factor F [Source:HGNC Symbol;Acc:HGNC:6780]	7.03
NR1H2	nuclear receptor subfamily 1 group H member 2 [Source:HGNC Symbol;Acc:HGNC:7965]	1.55
NR2C2	nuclear receptor subfamily 2 group C member 2 [Source:HGNC Symbol;Acc:HGNC:7972]	1.39
NR4A1	nuclear receptor subfamily 4 group A member 1 [Source:HGNC Symbol;Acc:HGNC:7980]	1.99
RXRB	retinoid X receptor beta [Source:HGNC Symbol;Acc:HGNC:10478]	2.24
SIRT6	sirtuin 6 [Source:HGNC Symbol;Acc:HGNC:14934]	1.25

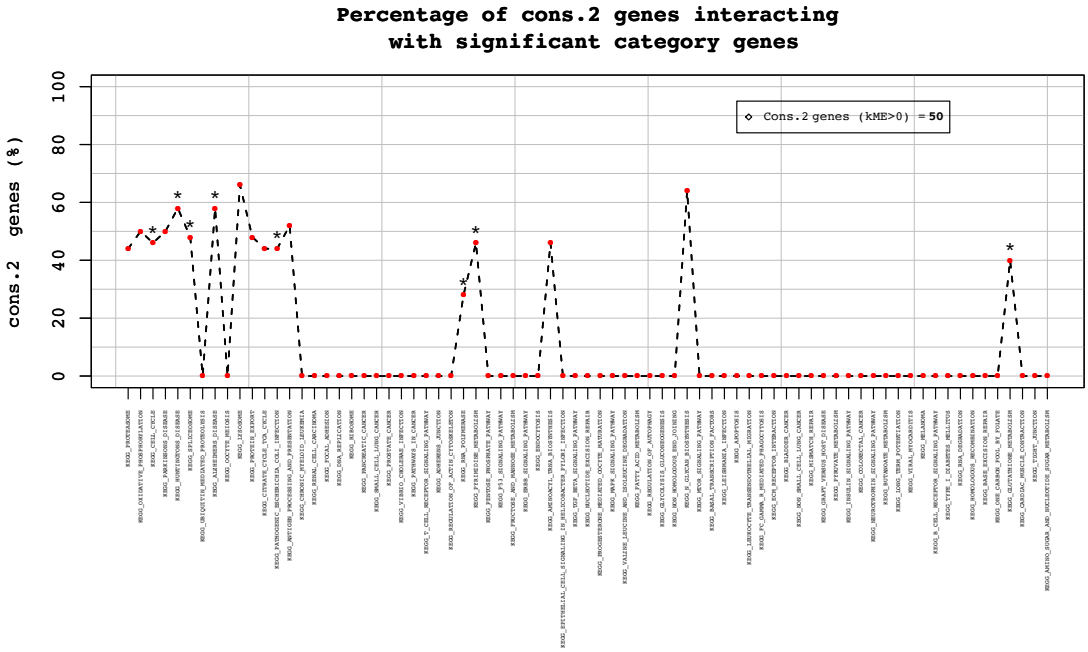
ZNF589	zinc finger protein 589 [Source:HGNC Symbol;Acc:HGNC:16747]	2.56
ZNF784	zinc finger protein 784 [Source:HGNC Symbol;Acc:HGNC:33111]	2.16
Cons.20		
CREB3L2	cAMP responsive element binding protein 3 like 2 [Source:HGNC Symbol;Acc:HGNC:23720]	2.12
EPAS1	endothelial PAS domain protein 1 [Source:HGNC Symbol;Acc:HGNC:3374]	7.00
GMEB2	glucocorticoid modulatory element binding protein 2 [Source:HGNC Symbol;Acc:HGNC:4371]	16.45
NFIC	nuclear factor I C [Source:HGNC Symbol;Acc:HGNC:7786]	2.44
POLR3A	RNA polymerase III subunit A [Source:HGNC Symbol;Acc:HGNC:30074]	1.52
STRA13	#N/A	2.28
ZBTB14	zinc finger and BTB domain containing 14 [Source:HGNC Symbol;Acc:HGNC:12860]	1.70
ZBTB18	zinc finger and BTB domain containing 18 [Source:HGNC Symbol;Acc:HGNC:13030]	2.43
ZNF282	zinc finger protein 282 [Source:HGNC Symbol;Acc:HGNC:13076]	1.92
ZNF652	zinc finger protein 652 [Source:HGNC Symbol;Acc:HGNC:29147]	1.68
Cons.22		
ARID5B	AT-rich interaction domain 5B [Source:HGNC Symbol;Acc:HGNC:17362]	1.28
ELK1	ELK1, ETS transcription factor [Source:HGNC Symbol;Acc:HGNC:3321]	1.04
ETS2	ETS proto-oncogene 2, transcription factor [Source:HGNC Symbol;Acc:HGNC:3489]	1.03
GMEB1	glucocorticoid modulatory element binding protein 1 [Source:HGNC Symbol;Acc:HGNC:4370]	1.24
MECP2	methyl-CpG binding protein 2 [Source:HGNC Symbol;Acc:HGNC:6990]	1.23
RFX3	regulatory factor X3 [Source:HGNC Symbol;Acc:HGNC:9984]	1.31
RFX5	regulatory factor X5 [Source:HGNC Symbol;Acc:HGNC:9986]	1.31
ZBTB14	zinc finger and BTB domain containing 14 [Source:HGNC Symbol;Acc:HGNC:12860]	1.30
ZBTB33	zinc finger and BTB domain containing 33 [Source:HGNC Symbol;Acc:HGNC:16682]	1.39
ZBTB7A	zinc finger and BTB domain containing 7A [Source:HGNC Symbol;Acc:HGNC:18078]	4.14
Cons.23		
CLOCK	clock circadian regulator [Source:HGNC Symbol;Acc:HGNC:2082]	1.91
EP300	E1A binding protein p300 [Source:HGNC Symbol;Acc:HGNC:3373]	2.85
GMEB2	glucocorticoid modulatory element binding protein 2 [Source:HGNC Symbol;Acc:HGNC:4371]	2.96
NFYA	nuclear transcription factor Y subunit alpha [Source:HGNC Symbol;Acc:HGNC:7804]	1.91
NFYC	nuclear transcription factor Y subunit gamma [Source:HGNC Symbol;Acc:HGNC:7806]	1.79
STAT5A	signal transducer and activator of transcription 5A [Source:HGNC	2.02

	Symbol;Acc:HGNC:11366]	
STAT5B	signal transducer and activator of transcription 5B [Source:HGNC Symbol;Acc:HGNC:11367]	1.87
TERF1	telomeric repeat binding factor 1 [Source:HGNC Symbol;Acc:HGNC:11728]	2.70
ZNF628	zinc finger protein 628 [Source:HGNC Symbol;Acc:HGNC:28054]	3.51
ZNF784	zinc finger protein 784 [Source:HGNC Symbol;Acc:HGNC:33111]	2.33
Cons.24		
DR1	down-regulator of transcription 1 [Source:HGNC Symbol;Acc:HGNC:3017]	1.79
ELK3	ELK3, ETS transcription factor [Source:HGNC Symbol;Acc:HGNC:3325]	1.51
MEF2D	myocyte enhancer factor 2D [Source:HGNC Symbol;Acc:HGNC:6997]	1.48
NR1H2	nuclear receptor subfamily 1 group H member 2 [Source:HGNC Symbol;Acc:HGNC:7965]	1.78
NR3C1	nuclear receptor subfamily 3 group C member 1 [Source:HGNC Symbol;Acc:HGNC:7978]	3.04
PTEN	phosphatase and tensin homolog [Source:HGNC Symbol;Acc:HGNC:9588]	2.07
RXRA	retinoid X receptor alpha [Source:HGNC Symbol;Acc:HGNC:10477]	1.52
RXRB	retinoid X receptor beta [Source:HGNC Symbol;Acc:HGNC:10478]	2.80
TFAP4	transcription factor AP-4 [Source:HGNC Symbol;Acc:HGNC:11745]	2.16
ZNF75A	zinc finger protein 75a [Source:HGNC Symbol;Acc:HGNC:13146]	1.65

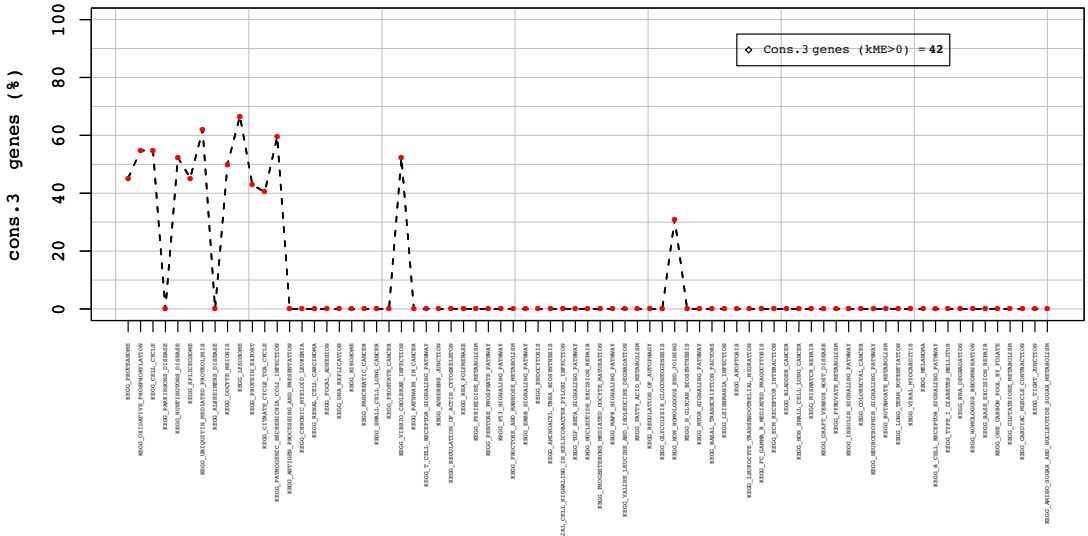
SUPPLEMENTARY FIGURES

Figures S1-S23 Percentage of module genes interacting with significant KEGG- or GO_MF-category genes. Red dots at the zero level indicate non-significant categories; by contrast, dots that are above zero indicate categories that are statistically significant (fdr<0.05) for the hypergeometric test. (*) denotes the categories that have statistical significance (fdr<0.05) for both hypergeometric test and DSEA; zoom in on the x-labels is necessary to read the categories.

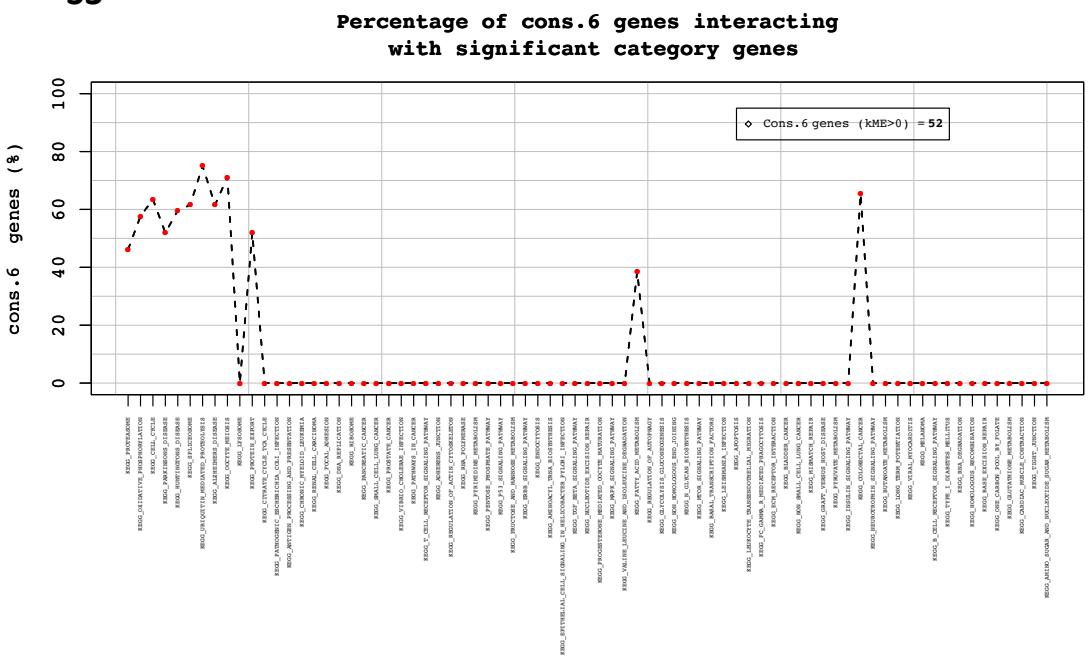
S1


Percentage of cons.2 genes interacting
with significant category genes
cons.2 genes (%)
Cons.2 genes (kME>0) = 50

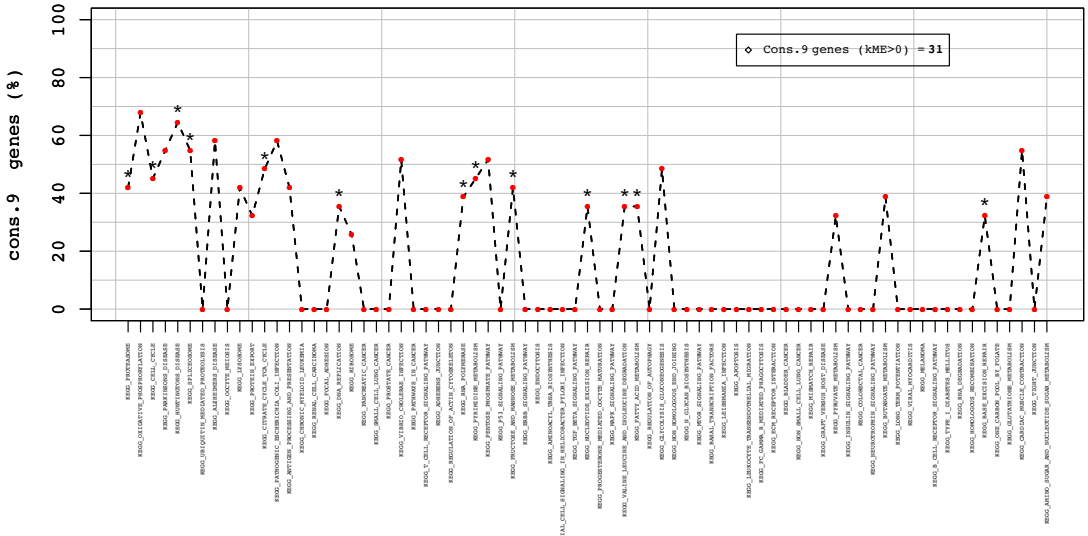
S2


Percentage of cons.3 genes interacting
with significant category genes
cons.3 genes (%)
Cons.3 genes (kME>0) = 42

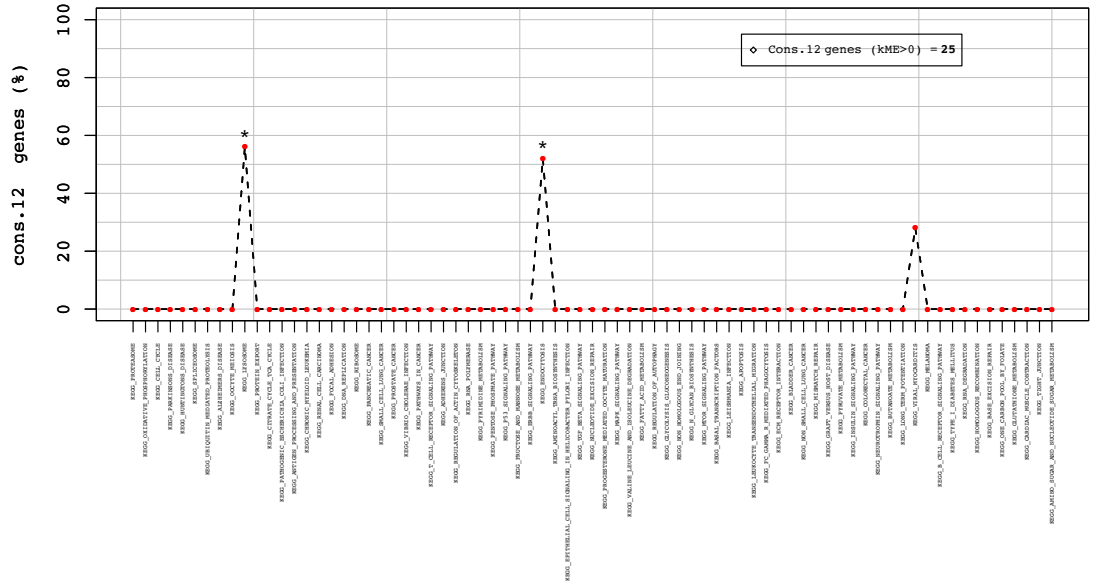
S3


Percentage of cons.6 genes interacting
with significant category genes
cons.6 genes (%)
Cons.6 genes (kME>0) = 52

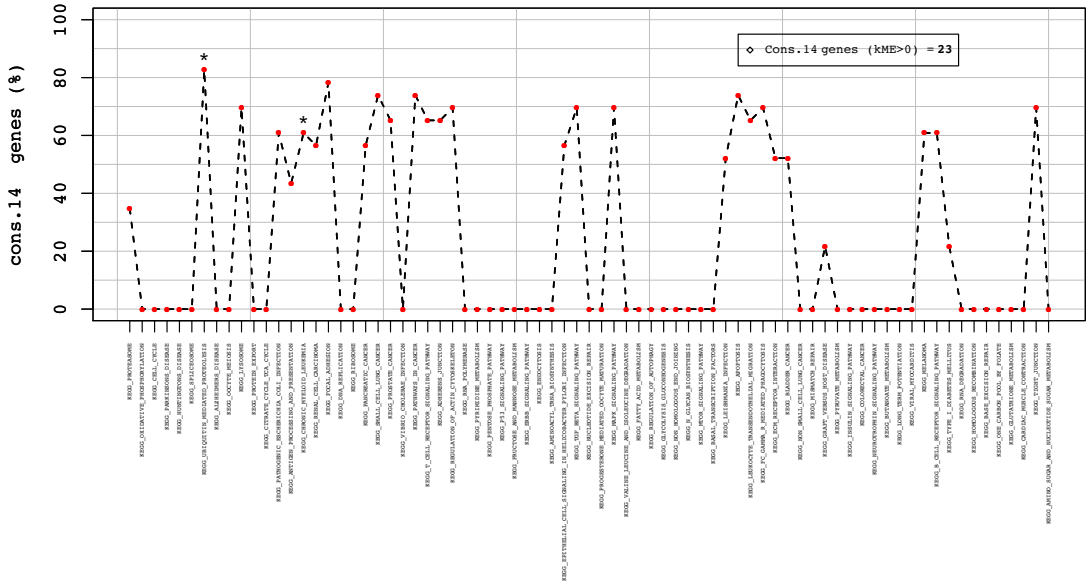
S4


Percentage of cons.9 genes interacting
with significant category genes
cons.9 genes (%)
Cons.9 genes (kME>0) = 31

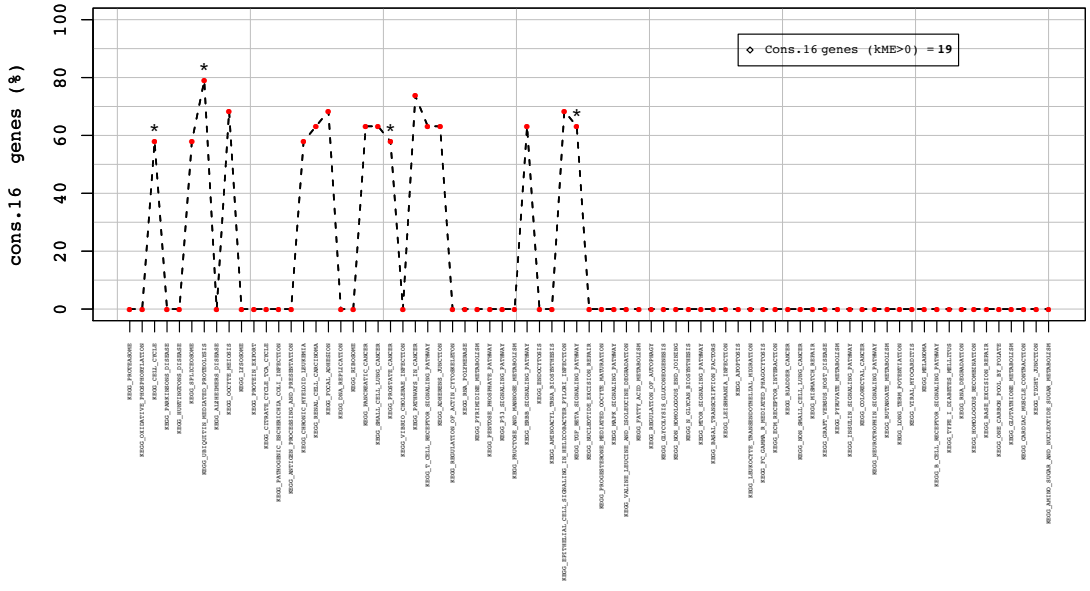
S5


Percentage of cons.12 genes interacting
with significant category genes
cons.12 genes (%)
Cons.12 genes (kME>0) = 25

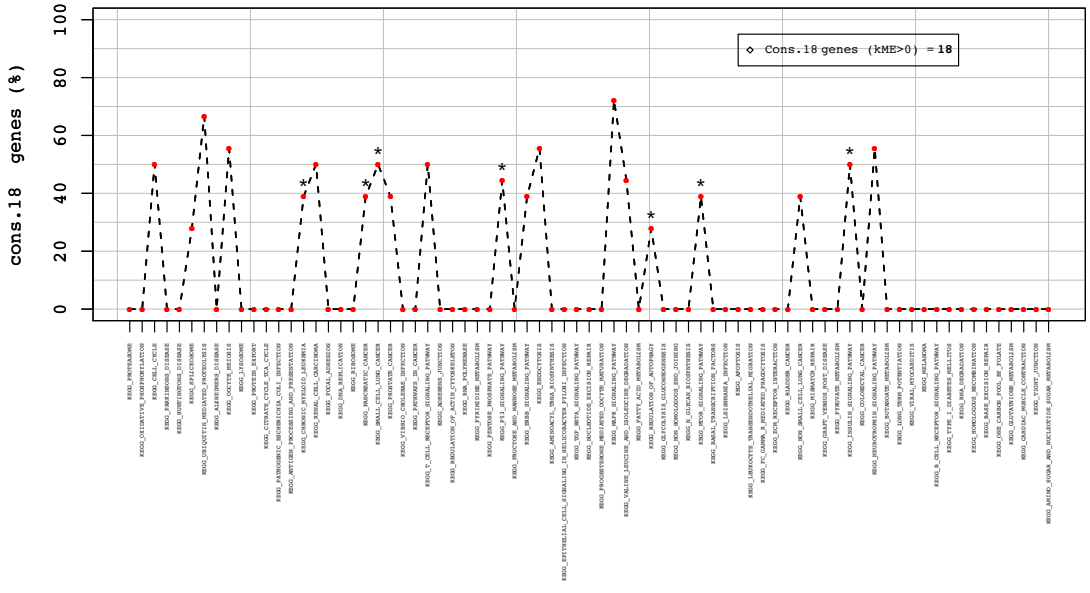
S6


Percentage of cons.14 genes interacting
with significant category genes
cons.14 genes (%)
Cons.14 genes (kME>0) = 23

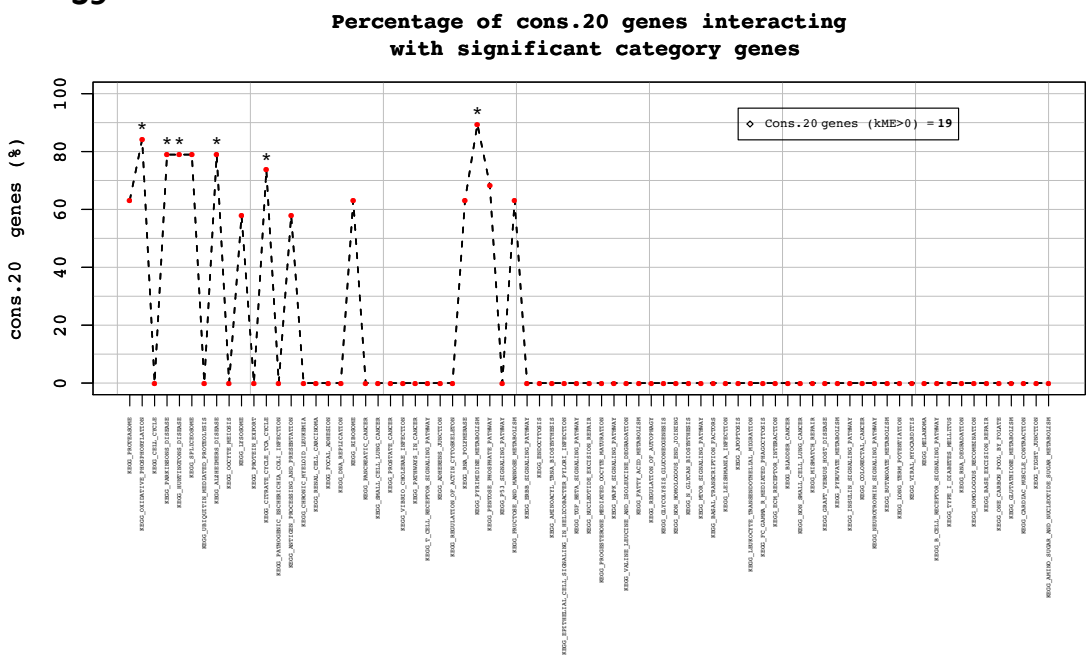
S7


Percentage of cons.16 genes interacting
with significant category genes
cons.16 genes (%)
Cons.16 genes (kME>0) = 19

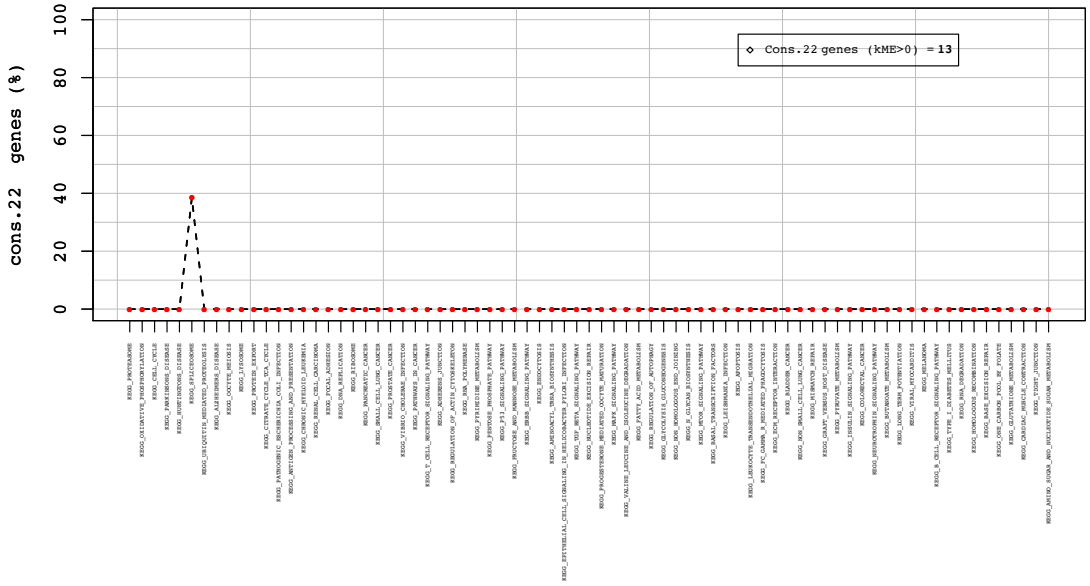
S8


Percentage of cons.18 genes interacting
with significant category genes
cons.18 genes (%)
Cons.18 genes (kME>0) = 18

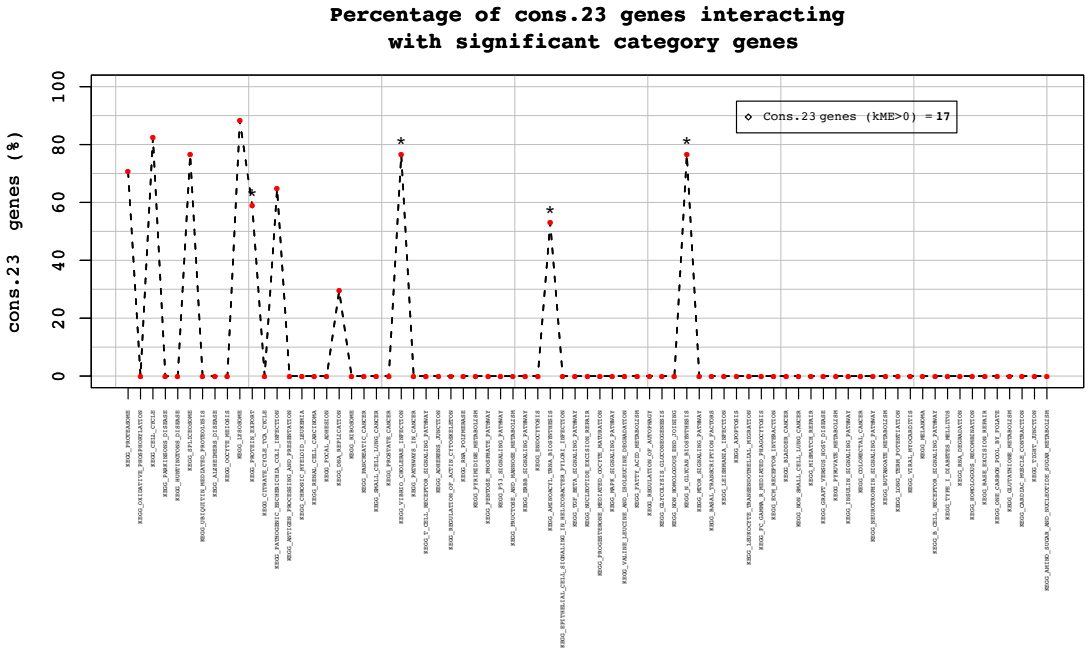
S9


Percentage of cons.20 genes interacting
with significant category genes
cons.20 genes (%)
Cons.20 genes (kME>0) = 19

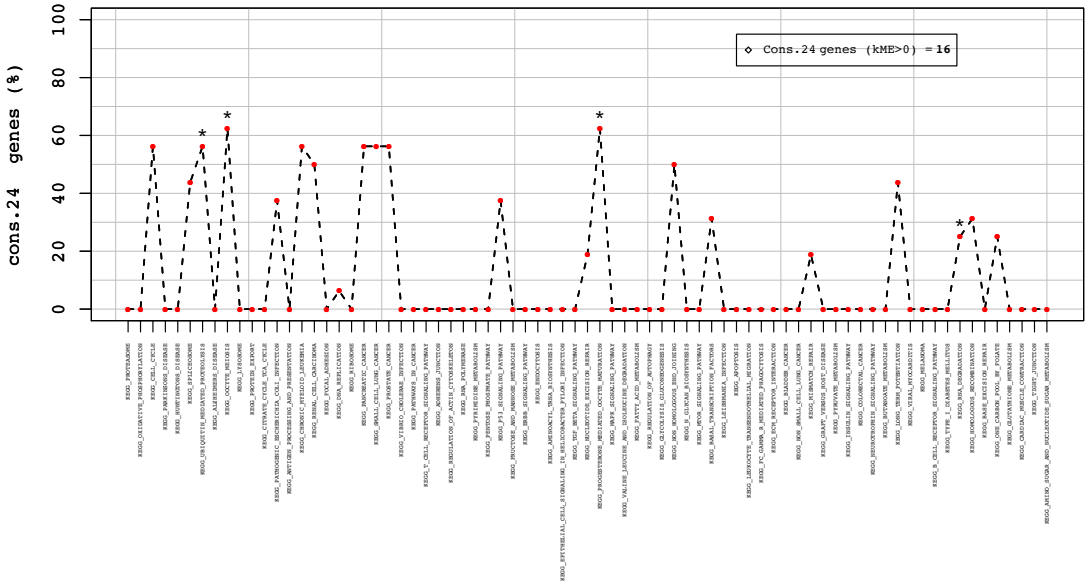
S10


Percentage of cons.22 genes interacting
with significant category genes
cons.22 genes (%)
Cons.22 genes (kME>0) = 13

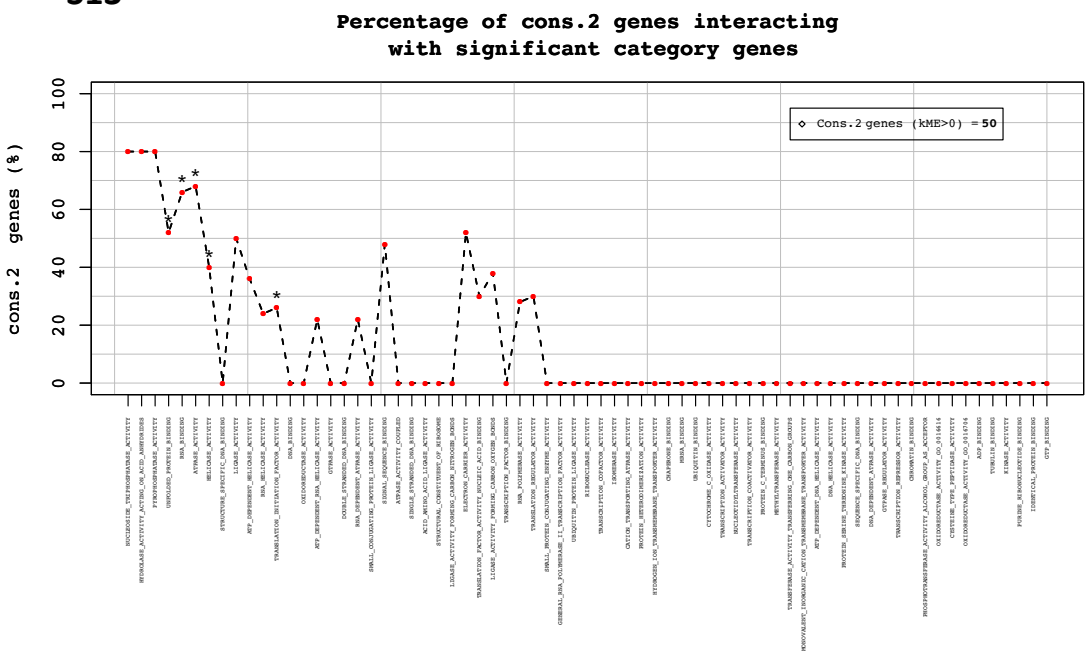
S11


Percentage of cons.23 genes interacting
with significant category genes
cons.23 genes (%)
Cons.23 genes (kME>0) = 17

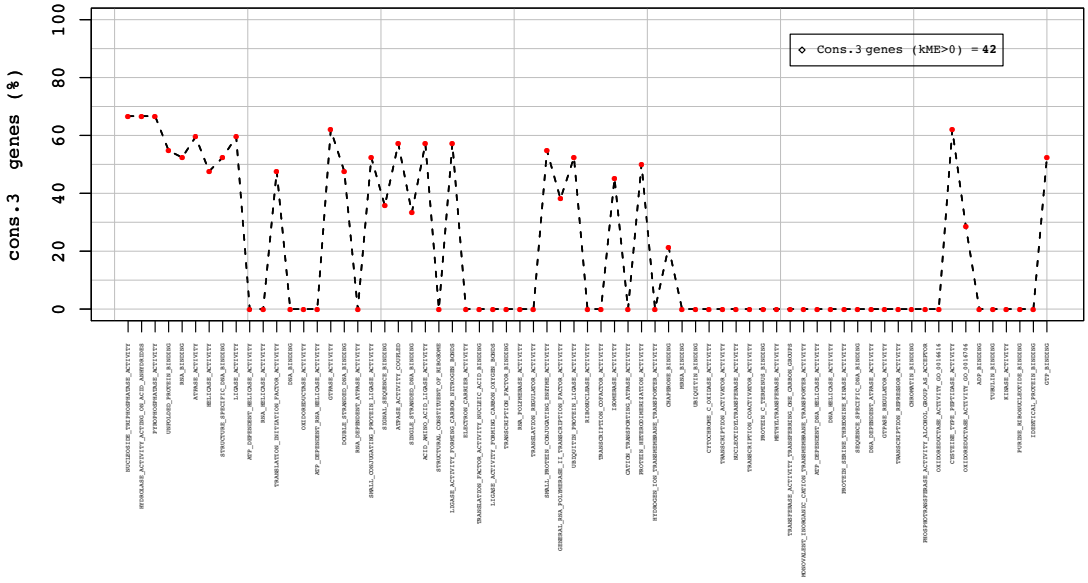
S12


Percentage of cons.24 genes interacting
with significant category genes
cons.24 genes (%)
Cons.24 genes (kME>0) = 16

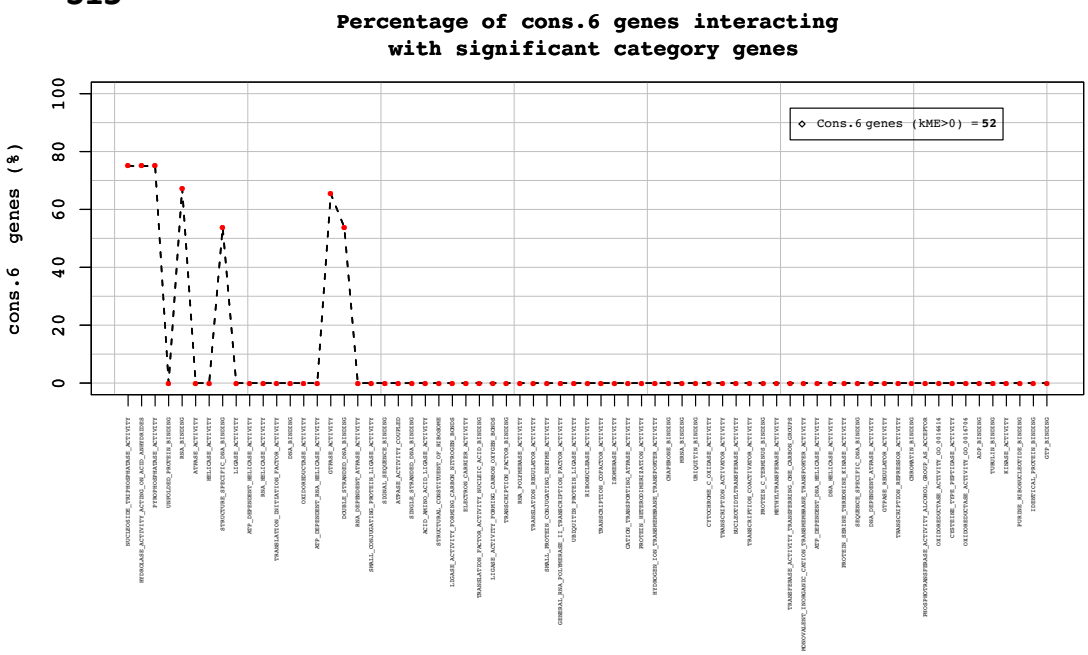
S13


Percentage of cons.2 genes interacting
with significant category genes
cons.2 genes (%)
Cons.2 genes (kME>0) = 50

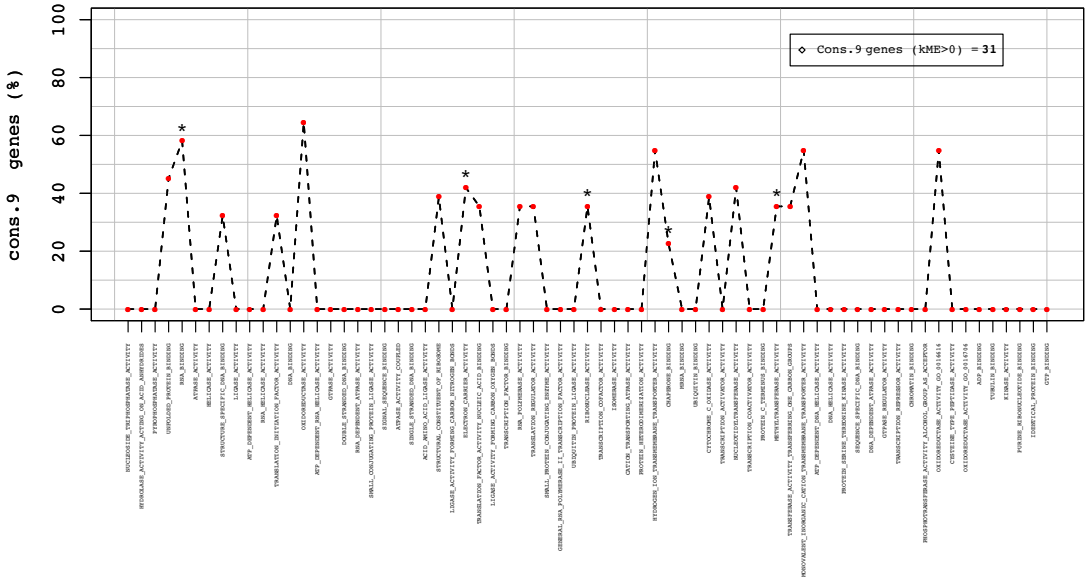
S14


Percentage of cons.3 genes interacting
with significant category genes
cons.3 genes (%)
Cons.3 genes (kME>0) = 42

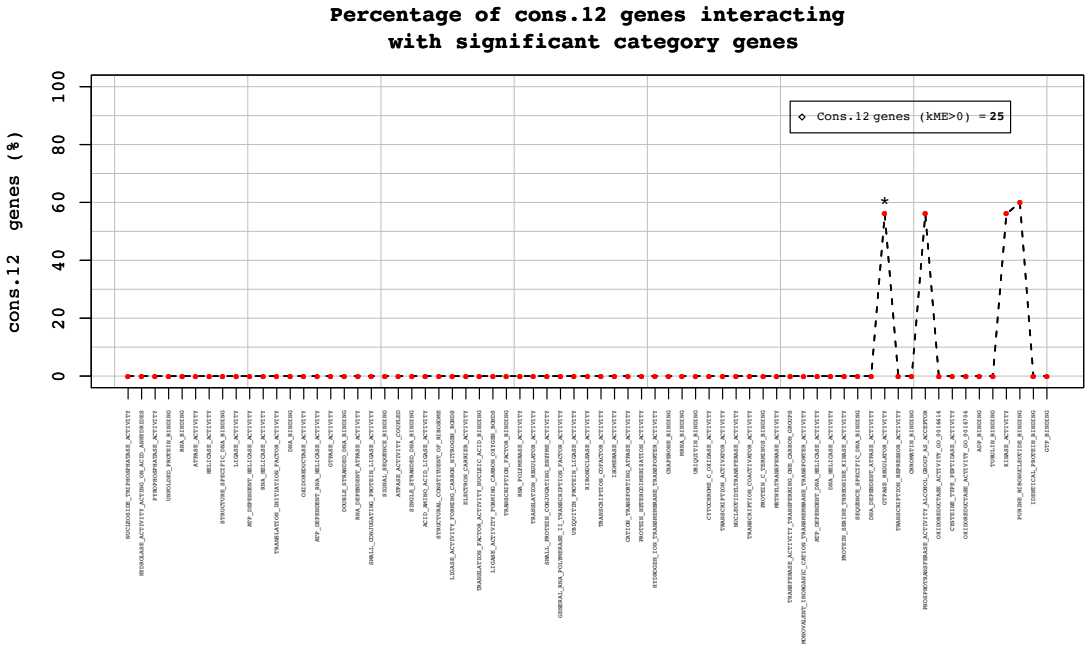
S15


Percentage of cons.6 genes interacting
with significant category genes
cons.6 genes (%)
0
20
40
60
80
100
Cons.6 genes (kME>0) = 52

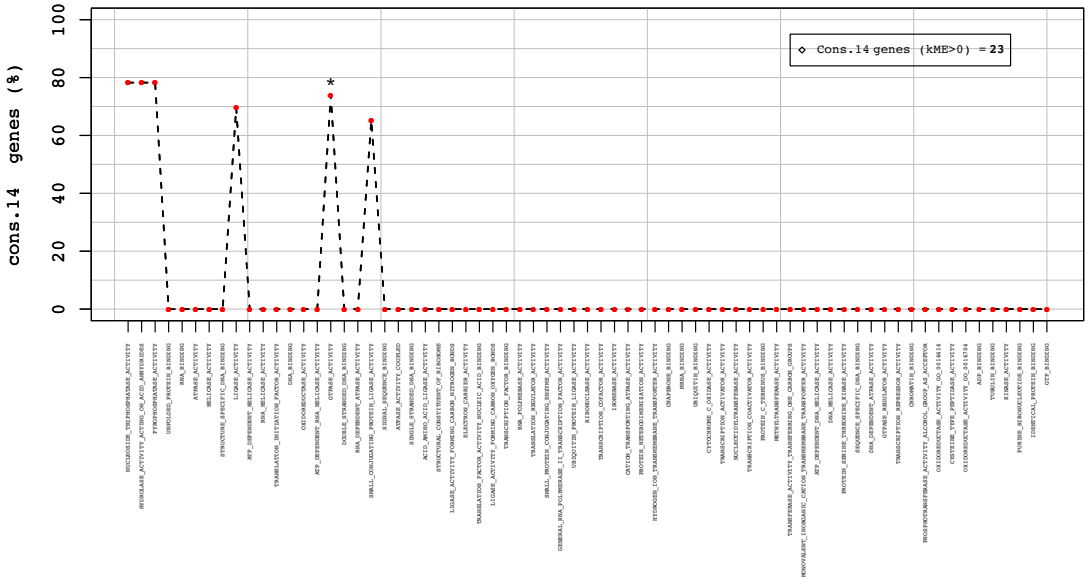
S16


Percentage of cons.9 genes interacting
with significant category genes
cons.9 genes (%)
0
20
40
60
80
100
Cons.9 genes (kME>0) = 31

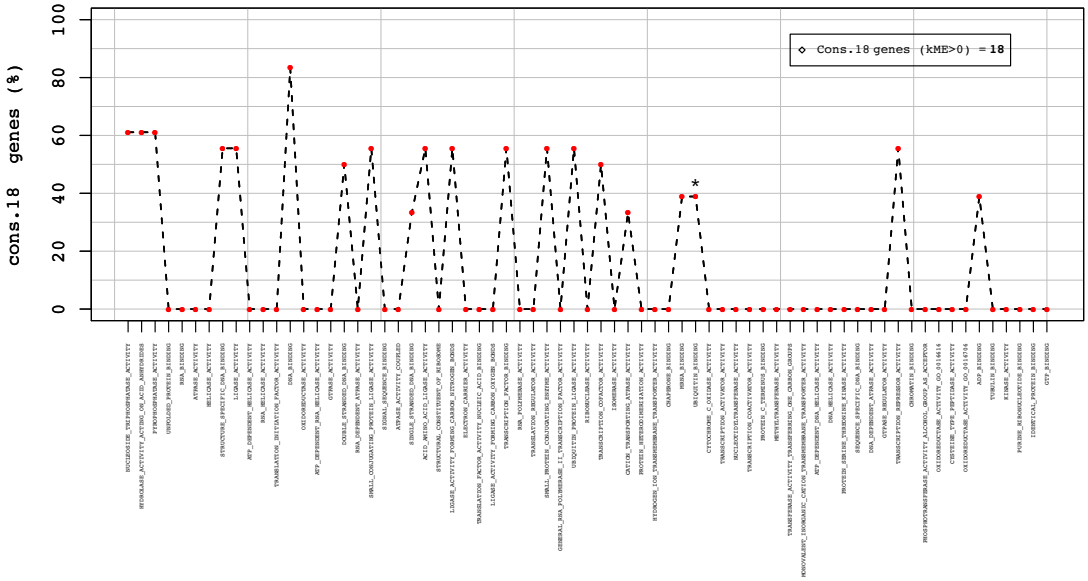
S17


Percentage of cons.12 genes interacting
with significant category genes
cons.12 genes (%)
Cons.12 genes (kME>0) = 25

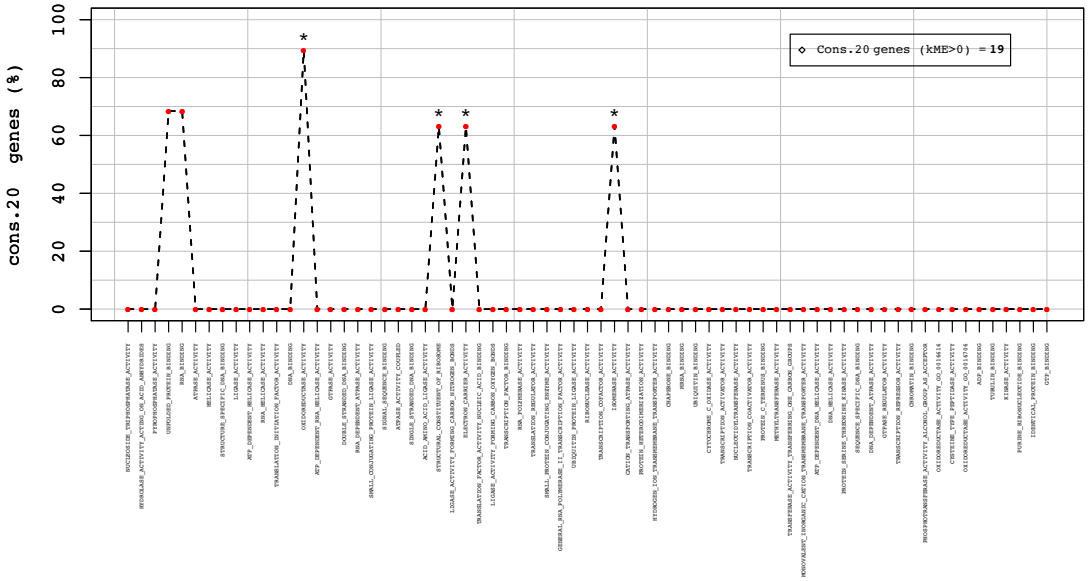
S18


Percentage of cons.14 genes interacting
with significant category genes
cons.14 genes (%)
Cons.14 genes (kME>0) = 23

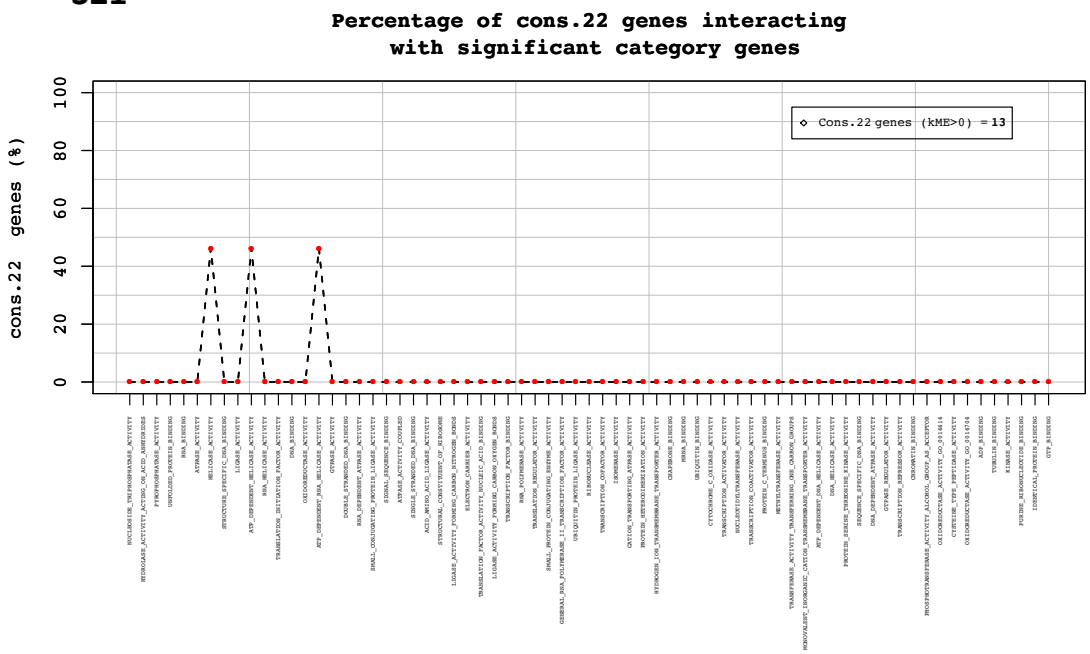
S19


Percentage of cons.18 genes interacting
with significant category genes
cons.18 genes (%)
0
20
40
60
80
100
Cons.18 genes (kME>0) = 18

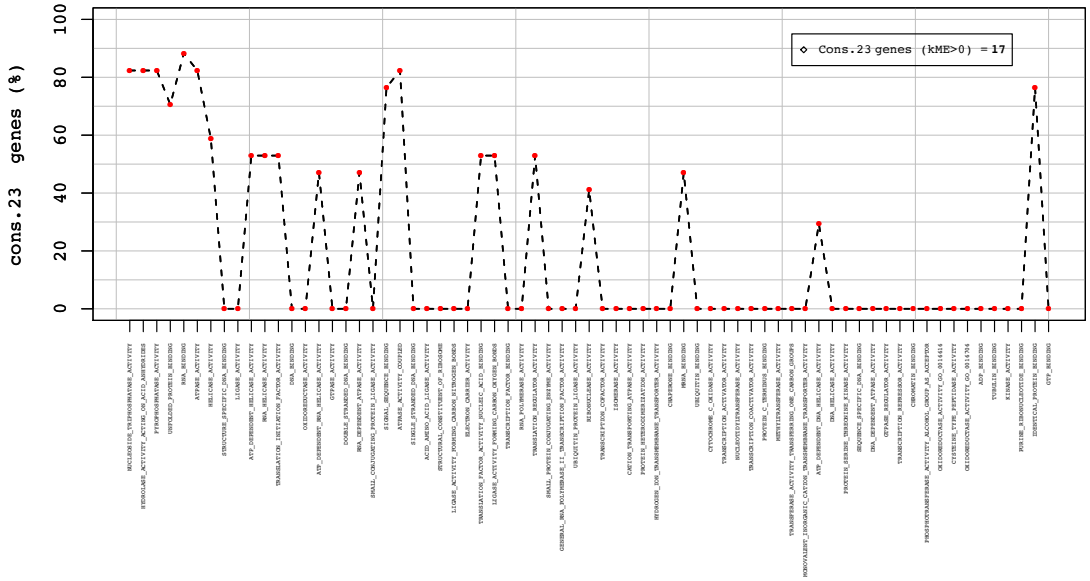
S20


Percentage of cons.20 genes interacting
with significant category genes
cons.20 genes (%)
0
20
40
60
80
100
Cons.20 genes (kME>0) = 19

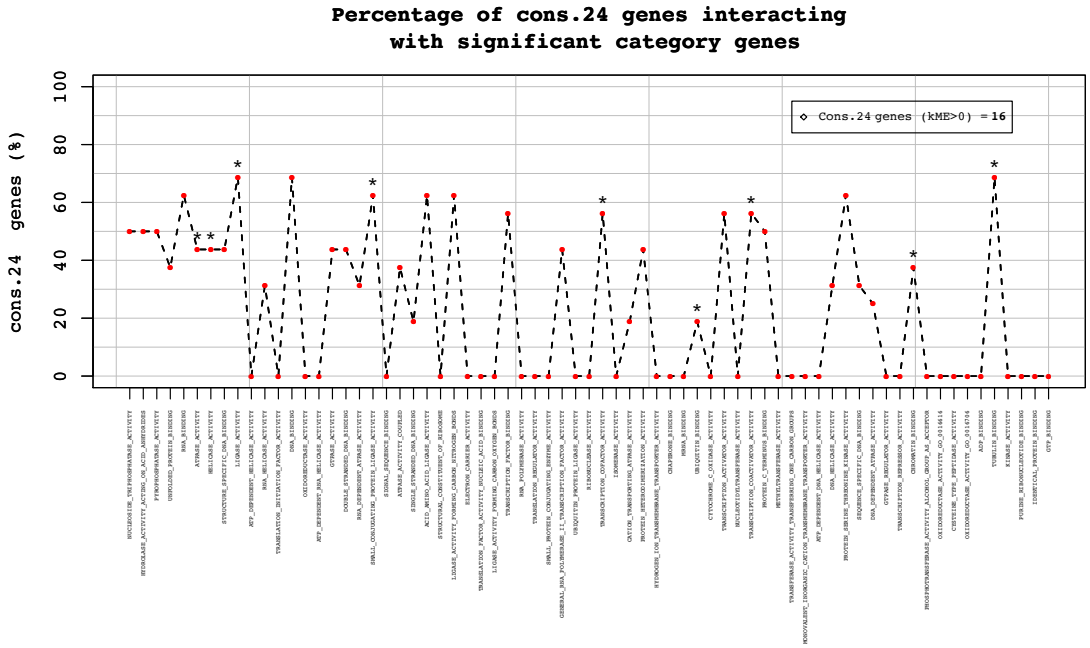
S21


Percentage of cons.22 genes interacting
with significant category genes
cons.22 genes (%)
Cons.22 genes (kME>0) = 13

S22


Percentage of cons.23 genes interacting
with significant category genes
cons.23 genes (%)
Cons.23 genes (kME>0) = 17

S23


Percentage of cons.24 genes interacting
with significant category genes
cons.24 genes (%)
0
20
40
60
80
100
Cons.24 genes (kME>0) = 16

Bibliography

Aït-Slimane, T., Galmes, R., Trugnan, G., and Maurice, M. (2009). Basolateral Internalization of GPI-anchored Proteins Occurs via a Clathrin-independent Flotillin-dependent Pathway in Polarized Hepatic Cells. Mol Biol Cell *20*, 3792–3800.

Amberger, J., Bocchini, C.A., Scott, A.F., and Hamosh, A. (2009). McKusick's Online Mendelian Inheritance in Man (OMIM®). Nucleic Acids Res *37*, D793–D796.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Arighi, C.N., Hartnell, L.M., Aguilar, R.C., Haft, C.R., and Bonifacino, J.S. (2004). Role of the mammalian retromer in sorting of the cation-independent mannose 6-phosphate receptor. J. Cell Biol. *165*, 123–133.

Babbey, C.M., Ahktar, N., Wang, E., Chen, C.C.-H., Grant, B.D., and Dunn, K.W. (2006). Rab10 regulates membrane transport through early endosomes of polarized Madin-Darby canine kidney cells. Mol. Biol. Cell *17*, 3156–3175.

Babst, M., Katzmann, D.J., Estepa-Sabal, E.J., Meerloo, T., and Emr, S.D. (2002a). Escrt-III: an endosome-associated heterooligomeric protein complex required for mvb sorting. Dev. Cell *3*, 271–282.

Babst, M., Katzmann, D.J., Snyder, W.B., Wendland, B., and Emr, S.D. (2002b). Endosome-associated complex, ESCRT-II, recruits transport machinery for protein sorting at the multivesicular body. Dev. Cell *3*, 283–289.

Bache, K.G., Raiborg, C., Mehlum, A., and Stenmark, H. (2003). STAM and Hrs are subunits of a multivalent ubiquitin-binding complex on early endosomes. J. Biol. Chem. *278*, 12513–12521.

Baker, R.W., Jeffrey, P.D., Zick, M., Phillips, B.P., Wickner, W.T., and Hughson, F.M. (2015). A direct role for the Sec1/Munc18-family protein Vps33 as a template for SNARE assembly. Science *349*, 1111–1114.

Barlowe, C., and Helenius, A. (2016). Cargo Capture and Bulk Flow in the Early Secretory Pathway. Annu. Rev. Cell Dev. Biol. *32*, 197–222.

Barlowe, C., and Schekman, R. (1993). SEC12 encodes a guanine-nucleotide-exchange factor essential for transport vesicle budding from the ER. Nature *365*, 347–349.

Barlowe, C.K., and Miller, E.A. (2013). Secretory protein biogenesis and traffic in the early secretory pathway. Genetics *193*, 383–410.

Barrowman, J., Sacher, M., and Ferro-Novick, S. (2000). TRAPP stably associates with the Golgi and is required for vesicle docking. EMBO J. *19*, 862–869.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A

Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) *57*, 289–300.

Bennett, M.K., Calakos, N., and Scheller, R.H. (1992). Syntaxin: a synaptic protein implicated in docking of synaptic vesicles at presynaptic active zones. Science *257*, 255–259.

Bi, X., Corpina, R.A., and Goldberg, J. (2002). Structure of the Sec23/24-Sar1 pre-budding complex of the COPII vesicle coat. Nature *419*, 271–277.

Bland, J.M., and Altman, D.G. (1995). Multiple significance tests: the Bonferroni method. BMJ *310*, 170.

Blatti, C., and Sinha, S. (2016). Characterizing gene sets using discriminative random walks with restart on heterogeneous biological networks. Bioinformatics *32*, 2167–2175.

Böhni, P.C., Deshaies, R.J., and Schekman, R.W. (1988). SEC11 is required for signal peptide processing and yeast cell growth. J. Cell Biol. *106*, 1035–1042.

Bonifacino, J.S., and Hurley, J.H. (2008). Retromer. Curr. Opin. Cell Biol. *20*, 427–436.

Bonifacino, J.S., and Rojas, R. (2006). Retrograde transport from endosomes to the trans-Golgi network. Nat. Rev. Mol. Cell Biol. *7*, 568–579.

Boucrot, E., Ferreira, A.P.A., Almeida-Souza, L., Debard, S., Vallis, Y., Howard, G., Bertot, L., Sauvonnet, N., and McMahon, H.T. (2015). Endophilin marks and controls a clathrin-independent endocytic pathway. Nature *517*, 460–465.

Braakman, I., and Hebert, D.N. (2013). Protein Folding in the Endoplasmic Reticulum. Cold Spring Harb Perspect Biol *5*.

Burda, P., Padilla, S.M., Sarkar, S., and Emr, S.D. (2002). Retromer function in endosome-to-Golgi retrograde transport is regulated by the yeast Vps34 PtdIns 3-kinase. J. Cell. Sci. *115*, 3889–3900.

Butte, A.J., and Kohane, I.S. (2000). Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pac Symp Biocomput 418–429.

Cacchiarelli, D., Trapnell, C., Ziller, M.J., Soumillon, M., Cesana, M., Karnik, R., Donaghey, J., Smith, Z.D., Ratanasirintrawoot, S., Zhang, X., et al. (2015). Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of Induced Pluripotency. Cell *162*, 412–424.

Cai, H., Reinisch, K., and Ferro-Novick, S. (2007). Coats, tethers, Rabs, and SNAREs work together to mediate the intracellular destination of a transport vesicle. Dev. Cell *12*, 671–682.

Caplan, S., Naslavsky, N., Hartnell, L.M., Lodge, R., Polishchuk, R.S., Donaldson, J.G.,

and Bonifacino, J.S. (2002). A tubular EHD1-containing compartment involved in the recycling of major histocompatibility complex class I molecules to the plasma membrane. EMBO J. *21*, 2557–2567.

Carlton, J., Bujny, M., Peter, B.J., Oorschot, V.M.J., Rutherford, A., Mellor, H., Klumperman, J., McMahon, H.T., and Cullen, P.J. (2004). Sorting nexin-1 mediates tubular endosome-to-TGN transport through coincidence sensing of high-curvature membranes and 3-phosphoinositides. Curr. Biol. *14*, 1791–1800.

Carlton, J., Bujny, M., Rutherford, A., and Cullen, P. (2005a). Sorting nexins--unifying trends and new perspectives. Traffic *6*, 75–82.

Carlton, J.G., Bujny, M.V., Peter, B.J., Oorschot, V.M.J., Rutherford, A., Arkell, R.S., Klumperman, J., McMahon, H.T., and Cullen, P.J. (2005b). Sorting nexin-2 is associated with tubular elements of the early endosome, but is not essential for retromer-mediated endosome-to-TGN transport. J. Cell. Sci. *118*, 4527–4539.

Carter, S.L., Brechbühler, C.M., Griffin, M., and Bond, A.T. (2004). Gene co-expression network topology provides a framework for molecular characterization of cellular state. Bioinformatics *20*, 2242–2250.

Chen, Y.A., and Scheller, R.H. (2001). SNARE-mediated membrane fusion. Nat. Rev. Mol. Cell Biol. *2*, 98–106.

Chia, P.Z.C., and Gleeson, P.A. (2014). Membrane tethering. F1000Prime Rep *6*, 74.

Chirico, W.J., Waters, M.G., and Blobel, G. (1988). 70K heat shock related proteins stimulate protein translocation into microsomes. Nature *332*, 805–810.

Choudhury, A., Sharma, D.K., Marks, D.L., and Pagano, R.E. (2004). Elevated endosomal cholesterol levels in Niemann-Pick cells inhibit rab4 and perturb membrane recycling. Mol. Biol. Cell *15*, 4500–4511.

Christoforidis, S., Miaczynska, M., Ashman, K., Wilm, M., Zhao, L., Yip, S.C., Waterfield, M.D., Backer, J.M., and Zerial, M. (1999). Phosphatidylinositol-3-OH kinases are Rab5 effectors. Nat. Cell Biol. *1*, 249–252.

Chung, J., Torta, F., Masai, K., Lucast, L., Czapla, H., Tanner, L.B., Narayanaswamy, P., Wenk, M.R., Nakatsu, F., and De Camilli, P. (2015). PI4P/phosphatidylserine countertransport at ORP5- and ORP8-mediated ER–plasma membrane contacts. Science *349*, 428–432.

Cockcroft, S., and De Matteis, M.A. (2001). Inositol lipids as spatial regulators of membrane traffic. J. Membr. Biol. *180*, 187–194.

Collins, B.M. (2008). The structure and function of the retromer protein complex. Traffic *9*, 1811–1822.

Cozier, G.E., Carlton, J., McGregor, A.H., Gleeson, P.A., Teasdale, R.D., Mellor, H., and Cullen, P.J. (2002). The phox homology (PX) domain-dependent, 3-phosphoinositide-mediated association of sorting nexin-1 with an early sorting

endosomal compartment is required for its ability to regulate epidermal growth factor receptor degradation. J. Biol. Chem. *277*, 48730–48736.

Dancourt, J., and Barlowe, C. (2010a). Protein sorting receptors in the early secretory pathway. Annu. Rev. Biochem. *79*, 777–802.

Dancourt, J., and Barlowe, C. (2010b). Protein sorting receptors in the early secretory pathway. Annu. Rev. Biochem. *79*, 777–802.

Davidson, G.S., Wylie, B.N., and Boyack, K.W. (2001). Cluster Stability and the Use of Noise in Interpretation of Clustering. In Proceedings of the IEEE Symposium on Information Visualization 2001 (INFOVIS'01), (Washington, DC, USA: IEEE Computer Society), p. 23–.

De Matteis, M.A., and Godi, A. (2004). PI-loting membrane traffic. Nat. Cell Biol. *6*, 487–492.

De Matteis, M.A., and Luini, A. (2011). Mendelian disorders of membrane trafficking. N. Engl. J. Med. *365*, 927–938.

De Matteis, M.A., and Rega, L.R. (2015). Endoplasmic reticulum-Golgi complex membrane contact sites. Curr. Opin. Cell Biol. *35*, 43–50.

De Matteis, M.A., Vicinanza, M., Venditti, R., and Wilson, C. (2013). Cellular assays for drug discovery in genetic disorders of intracellular trafficking. Annu Rev Genomics Hum Genet *14*, 159–190.

Deneka, M., Neeft, M., Popa, I., van Oort, M., Sprong, H., Oorschot, V., Klumperman, J., Schu, P., and van der Sluijs, P. (2003). Rabaptin-5alpha/rabaptin-4 serves as a linker between rab4 and gamma(1)-adaptin in membrane recycling from endosomes. EMBO J. *22*, 2645–2657.

Deshaies, R.J., Koch, B.D., Werner-Washburne, M., Craig, E.A., and Schekman, R. (1988). A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides. Nature *332*, 800–805.

Di Paolo, G., and De Camilli, P. (2006). Phosphoinositides in cell regulation and membrane dynamics. Nature *443*, 651–657.

Doherty, G.J., and McMahon, H.T. (2009). Mechanisms of endocytosis. Annu. Rev. Biochem. *78*, 857–902.

Dulubova, I., Yamaguchi, T., Gao, Y., Min, S.-W., Huryeva, I., Südhof, T.C., and Rizo, J. (2002). How Tlg2p/syntaxin 16 "snares" Vps45. EMBO J. *21*, 3620–3631.

Edgar, R., Mazor, Y., Rinon, A., Blumenthal, J., Golan, Y., Buzhor, E., Livnat, I., Ben-Ari, S., Lieder, I., Shitrit, A., et al. (2013). LifeMap Discovery™: the embryonic development, stem cells, and regenerative medicine research portal. PLoS ONE *8*, e66629.

Espenshade, P., Gimeno, R.E., Holzmacher, E., Teung, P., and Kaiser, C.A. (1995).

Yeast SEC16 gene encodes a multidomain vesicle coat protein that interacts with Sec23p. J. Cell Biol. *131*, 311–324.

Fisher, R.A. (1922). On the Interpretation of χ 2 from Contingency Tables and the Calculation of P.

Fotin, A., Cheng, Y., Sliz, P., Grigorieff, N., Harrison, S.C., Kirchhausen, T., and Walz, T. (2004). Molecular model for a complete clathrin lattice from electron cryomicroscopy. Nature *432*, 573–579.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., et al. (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell *153*, 1149–1163.

Gilbert, S.F. (2000). Paraxial Mesoderm: The Somites and Their Derivatives.

Godi, A., Pertile, P., Meyers, R., Marra, P., Di Tullio, G., Iurisci, C., Luini, A., Corda, D., and De Matteis, M.A. (1999). ARF mediates recruitment of PtdIns-4-OH kinase-beta and stimulates synthesis of PtdIns(4,5)P2 on the Golgi complex. Nat. Cell Biol. *1*, 280–287.

Godi, A., Di Campli, A., Konstantakopoulos, A., Di Tullio, G., Alessi, D.R., Kular, G.S., Daniele, T., Marra, P., Lucocq, J.M., and De Matteis, M.A. (2004). FAPPs control Golgi-to-cell-surface membrane traffic by binding to ARF and PtdIns(4)P. Nat. Cell Biol. *6*, 393–404.

Gomez-Navarro, N., and Miller, E. (2016). Protein sorting at the ER-Golgi interface. J. Cell Biol. *215*, 769–778.

Gramatica, R., Di Matteo, T., Giorgetti, S., Barbiani, M., Bevec, D., and Aste, T. (2014). Graph theory enables drug repurposing--how a mathematical model can drive the discovery of hidden mechanisms of action. PLoS ONE *9*, e84912.

Grant, B.D., and Donaldson, J.G. (2009). Pathways and mechanisms of endocytic recycling. Nature Reviews Molecular Cell Biology *10*, nrm2755.

Greene, C.S., Krishnan, A., Wong, A.K., Ricciotti, E., Zelaya, R.A., Himmelstein, D.S., Zhang, R., Hartmann, B.M., Zaslavsky, E., Sealfon, S.C., et al. (2015). Understanding multicellular function and disease with human tissue-specific networks. Nat. Genet. *47*, 569–576.

Gruenberg, J. (2001). The endocytic pathway: a mosaic of domains. Nat. Rev. Mol. Cell Biol. *2*, 721–730.

GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.

GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660.

Haft, C.R., de la Luz Sierra, M., Bafford, R., Lesniak, M.A., Barr, V.A., and Taylor, S.I. (2000). Human orthologs of yeast vacuolar protein sorting proteins Vps26, 29, and 35: assembly into multimeric complexes. Mol. Biol. Cell *11*, 4105–4116.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

Hattula, K., Furuhjelm, J., Tikkanen, J., Tanhuanpää, K., Laakkonen, P., and Peränen, J. (2006). Characterization of the Rab8-specific membrane traffic route linked to protrusion formation. J. Cell. Sci. *119*, 4866–4877.

Heider, D., Hauke, S., Pyka, M., and Kessler, D. (2010). Insights into the classification of small GTPases. Adv Appl Bioinforma Chem *3*, 15–24.

Helenius, A., and Aebi, M. (2004). Roles of N-linked glycans in the endoplasmic reticulum. Annu. Rev. Biochem. *73*, 1019–1049.

Hierro, A., Rojas, A.L., Rojas, R., Murthy, N., Effantin, G., Kajava, A.V., Steven, A.C., Bonifacino, J.S., and Hurley, J.H. (2007). Functional architecture of the retromer cargo-recognition complex. Nature *449*, 1063–1067.

Hofree, M., Shen, J.P., Carter, H., Gross, A., and Ideker, T. (2013). Network-based stratification of tumor mutations. Nat Methods *10*, 1108–1115.

Holthuis, J.C.M., and Levine, T.P. (2005). Lipid traffic: floppy drives and a superhighway. Nat. Rev. Mol. Cell Biol. *6*, 209–220.

Hopkins, C.R., and Trowbridge, I.S. (1983). Internalization and processing of transferrin and the transferrin receptor in human carcinoma A431 cells. J. Cell Biol. *97*, 508–521.

Horvath, S. (2011). Weighted Network Analysis: Applications in Genomics and Systems Biology (Springer Science & Business Media).

Hou, J.P., and Ma, J. (2014). DawnRank: discovering personalized driver genes in cancer. Genome Med *6*.

Hu, J.X., Thomas, C.E., and Brunak, S. (2016). Network biology concepts in complex disease comorbidities. Nat. Rev. Genet. *17*, 615–629.

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., et al. (2002). The Ensembl genome database project. Nucleic Acids Res. *30*, 38–41.

Hurley, J.H. (2010). The ESCRT complexes. Crit. Rev. Biochem. Mol. Biol. *45*, 463–487.

Inoue, H., Ha, V.L., Prekeris, R., and Randazzo, P.A. (2008). Arf GTPase-activating Protein ASAP1 Interacts with Rab11 Effector FIP3 and Regulates Pericentrosomal Localization of Transferrin Receptor–positive Recycling Endosome. Mol Biol Cell

19, 4224–4237.

Iorio, F., Bosotti, R., Scacheri, E., Belcastro, V., Mithbaokar, P., Ferriero, R., Murino, L., Tagliaferri, R., Brunetti-Pierri, N., Isacchi, A., et al. (2010). Discovery of drug mode of action and drug repositioning from transcriptional responses. Proc. Natl. Acad. Sci. U.S.A. *107*, 14621–14626.

Ishikawa, T., Toyama, T., Nakamura, Y., Tamada, K., Shimizu, H., Ninagawa, S., Okada, T., Kamei, Y., Ishikawa-Fujiwara, T., Todo, T., et al. (2017). UPR transducer BBF2H7 allows export of type II collagen in a cargo- and developmental stage-specific manner. J. Cell Biol. *216*, 1761–1774.

Iván, G., and Grolmusz, V. (2011). When the Web meets the cell: using personalized PageRank for analyzing protein interaction networks. Bioinformatics *27*, 405–407.

Jakob, C.A., Burda, P., Roth, J., and Aebi, M. (1998). Degradation of misfolded endoplasmic reticulum glycoproteins in Saccharomyces cerevisiae is determined by a specific oligosaccharide structure. J. Cell Biol. *142*, 1223–1233.

Jaskowiak, P.A., Campello, R.J.G.B., and Costa, I.G. (2014). On the selection of appropriate distances for gene expression data clustering. BMC Bioinformatics *15 Suppl 2*, S2.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. Cell *152*, 327–339.

Jovic, M., Sharma, M., Rahajeng, J., and Caplan, S. (2010). The early endosome: a busy sorting station for proteins at the crossroads. Histol. Histopathol. *25*, 99–112.

Kalaitzis, A.A., and Lawrence, N.D. (2011). A simple approach to ranking differentially expressed gene expression time courses through Gaussian process regression. BMC Bioinformatics *12*, 180.

Kelleher, D.J., and Gilmore, R. (2006). An evolving view of the eukaryotic oligosaccharyltransferase. Glycobiology *16*, 47R–62R.

Kerr, M.C., and Teasdale, R.D. (2009). Defining macropinocytosis. Traffic *10*, 364–371.

Kim, J.J., Lipatova, Z., and Segev, N. (2016). TRAPP Complexes in Secretion and Autophagy. Front Cell Dev Biol *4*, 20.

Kornfeld, R., and Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. Annu. Rev. Biochem. *54*, 631–664.

Kouranti, I., Sachse, M., Arouche, N., Goud, B., and Echard, A. (2006). Rab35 regulates an endocytic recycling pathway essential for the terminal steps of cytokinesis. Curr. Biol. *16*, 1719–1725.

Kumari, S., and Mayor, S. (2008). ARF1 is directly involved in dynamin-

independent endocytosis. Nat. Cell Biol. *10*, 30–41.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics *24*, 719–720.

Langfelder, P., Luo, R., Oldham, M.C., and Horvath, S. (2011). Is My Network Module Preserved and Reproducible? PLoS Comput Biol *7*.

Lev, S. (2010). Non-vesicular lipid transport by lipid-transfer proteins and beyond. Nat. Rev. Mol. Cell Biol. *11*, 739–750.

Li, L., Wang, Y., An, L., Kong, X., and Huang, T. (2017). A network-based method using a random walk with restart algorithm and screening tests to identify novel genes associated with Menière's disease. PLoS One *12*.

Liberali, P., Snijder, B., and Pelkmans, L. (2014). A hierarchical map of regulatory genetic interactions in membrane trafficking. Cell *157*, 1473–1487.

Lin, Y., Golovnina, K., Chen, Z.-X., Lee, H.N., Negron, Y.L.S., Sultana, H., Oliver, B., and Harbison, S.T. (2016). Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual Drosophila melanogaster. BMC Genomics *17*.

Lobingier, B.T., Nickerson, D.P., Lo, S.-Y., and Merz, A.J. (2014). SM proteins Sly1 and Vps33 co-assemble with Sec17 and SNARE complexes to oppose SNARE disassembly by Sec18. Elife *3*, e02272.

Losev, E., Reinke, C.A., Jellen, J., Strongin, D.E., Bevis, B.J., and Glick, B.S. (2006). Golgi maturation visualized in living yeast. Nature *441*, 1002–1006.

Lovász, L. (1993). Random Walks on Graphs: A Survey, Combinatorics, Paul Erdos is Eighty. Bolyai Soc. Math. Stud. *2*.

Macqueen, J. (1967). Some methods for classification and analysis of multivariate observations. In In 5-Th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297.

Magadán, J.G., Barbieri, M.A., Mesa, R., Stahl, P.D., and Mayorga, L.S. (2006). Rab22a regulates the sorting of transferrin to recycling endosomes. Mol. Cell. Biol. *26*, 2595–2614.

Manford, A.G., Stefan, C.J., Yuan, H.L., Macgurn, J.A., and Emr, S.D. (2012). ER-to-plasma membrane tethering proteins regulate cell signaling and ER morphology. Dev. Cell *23*, 1129–1140.

Mann, H.B., and Whitney, D.R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. The Annals of Mathematical

Statistics *18*, 50–60.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. *18*, 1509–1517.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C., Chou, A., Ienasescu, H., et al. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res. *42*, D142-147.

Matsuura-Tokita, K., Takeuchi, M., Ichihara, A., Mikuriya, K., and Nakano, A. (2006). Live imaging of yeast Golgi cisternal maturation. Nature *441*, 1007–1010.

Mattera, R., and Bonifacino, J.S. (2008). Ubiquitin binding and conjugation regulate the recruitment of Rabex-5 to early endosomes. EMBO J. *27*, 2484–2494.

Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., et al. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. *34*, D108-110.

Mayor, S., Presley, J.F., and Maxfield, F.R. (1993). Sorting of membrane components from endosomes and subsequent recycling to the cell surface occurs by a bulk flow process. J. Cell Biol. *121*, 1257–1269.

McNew, J.A., Sogaard, M., Lampen, N.M., Machida, S., Ye, R.R., Lacomis, L., Tempst, P., Rothman, J.E., and Söllner, T.H. (1997). Ykt6p, a prenylated SNARE essential for endoplasmic reticulum-Golgi transport. J. Biol. Chem. *272*, 17776–17783.

Mellman, I. (1996). Endocytosis and molecular sorting. Annu. Rev. Cell Dev. Biol. *12*, 575–625.

von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. (2003). STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. *31*, 258–261.

Mesmin, B., Bigay, J., Moser von Filseck, J., Lacas-Gervais, S., Drin, G., and Antonny, B. (2013). A four-step cycle driven by PI(4)P hydrolysis directs sterol/PI(4)P exchange by the ER-Golgi tether OSBP. Cell *155*, 830–843.

Michell, R.H. (2011). Inositol and its derivatives: their evolution and functions. Adv. Enzyme Regul. *51*, 84–90.

Molendijk, A.J., Ruperti, B., and Palme, K. (2004). Small GTPases in vesicle trafficking. Curr Opin Plant Biol *7*, 694–700.

Murray, J.T., Panaretou, C., Stenmark, H., Miaczynska, M., and Backer, J.M. (2002). Role of Rab5 in the recruitment of hVps34/p150 to the early endosome. Traffic *3*, 416–427.

Napolitano, F., Sirci, F., Carrella, D., and di Bernardo, D. (2016). Drug-set enrichment analysis: a novel tool to investigate drug mode of action. Bioinformatics *32*, 235–241.

Naslavsky, N., Boehm, M., Backlund, P.S., and Caplan, S. (2004a). Rabenosyn-5 and EHD1 Interact and Sequentially Regulate Protein Recycling to the Plasma Membrane. Mol Biol Cell *15*, 2410–2422.

Naslavsky, N., Weigert, R., and Donaldson, J.G. (2004b). Characterization of a nonclathrin endocytic pathway: membrane cargo and lipid requirements. Mol. Biol. Cell *15*, 3542–3552.

Naslavsky, N., Rahajeng, J., Sharma, M., Jovic, M., and Caplan, S. (2006). Interactions between EHD proteins and Rab11-FIP2: a role for EHD3 in early endosomal transport. Mol. Biol. Cell *17*, 163–177.

Neph, S., Stergachis, A.B., Reynolds, A., Sandstrom, R., Borenstein, E., and Stamatoyannopoulos, J.A. (2012). Circuitry and dynamics of human transcription factor regulatory networks. Cell *150*, 1274–1286.

Newburger, D.E., and Bulyk, M.L. (2009). UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. Nucleic Acids Res. *37*, D77-82.

Niedergang, F., and Chavrier, P. (2004). Signaling and membrane dynamics during phagocytosis: many roads lead to the phagos(R)ome. Curr. Opin. Cell Biol. *16*, 422–428.

Nothwehr, S.F., Ha, S.A., and Bruinsma, P. (2000). Sorting of yeast membrane proteins into an endosome-to-Golgi pathway involves direct interaction of their cytosolic domains with Vps35p. J. Cell Biol. *151*, 297–310.

Ogg, S.C., Barz, W.P., and Walter, P. (1998). A functional GTPase domain, but not its transmembrane domain, is required for function of the SRP receptor beta-subunit., A Functional GTPase Domain, but not its Transmembrane Domain, is Required for Function of the SRP Receptor β-subunit. J Cell Biol *142*, *142*, 341, 341–354.

Orlean, P., and Menon, A.K. (2007). Thematic review series: lipid posttranslational modifications. GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycophospholipids. J. Lipid Res. *48*, 993–1011.

Orth, J.D., Krueger, E.W., Weller, S.G., and McNiven, M.A. (2006). A novel endocytic mechanism of epidermal growth factor receptor sequestration and internalization. Cancer Res. *66*, 3603–3610.

Pagano, A., Crottet, P., Prescianotto-Baschong, C., and Spiess, M. (2004). In vitro formation of recycling vesicles from endosomes requires adaptor protein-1/clathrin and is regulated by rab4 and the connector rabaptin-5. Mol. Biol. Cell *15*, 4990–5000.

Parton, R.G., Hanzal-Bayer, M., and Hancock, J.F. (2006). Biogenesis of caveolae: a

structural model for caveolin-induced domain formation. J. Cell. Sci. *119*, 787–796.

Payne, C.K., Jones, S.A., Chen, C., and Zhuang, X. (2007). Internalization and trafficking of cell surface proteoglycans and proteoglycan-binding ligands. Traffic *8*, 389–401.

Penfold, C.A., and Wild, D.L. (2011). How to infer gene networks from expression profiles, revisited. Interface Focus *1*, 857–870.

Plaisier, C.L., O'Brien, S., Bernard, B., Reynolds, S., Simon, Z., Toledo, C.M., Ding, Y., Reiss, D.J., Paddison, P.J., and Baliga, N.S. (2016). Causal Mechanistic Regulatory Network for Glioblastoma Deciphered Using Systems Genetics Network Analysis. Cell Syst *3*, 172–186.

Raiborg, C., Bache, K.G., Gillooly, D.J., Madshus, I.H., Stang, E., and Stenmark, H. (2002). Hrs sorts ubiquitinated proteins into clathrin-coated microdomains of early endosomes. Nat. Cell Biol. *4*, 394–398.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D., and Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol *14*, 3158.

Rapoport, T.A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature *450*, 663–669.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabási, A.L. (2002). Hierarchical organization of modularity in metabolic networks. Science *297*, 1551–1555.

Redecker, P. (1996). Synaptotagmin I, synaptobrevin II, and syntaxin I are coexpressed in rat and gerbil pinealocytes. Cell Tissue Res. *283*, 443–454.

Reiling, J.H., Olive, A.J., Sanyal, S., Carette, J.E., Brummelkamp, T.R., Ploegh, H.L., Starnbach, M.N., and Sabatini, D.M. (2013). A CREB3-ARF4 signalling pathway mediates the response to Golgi stress and susceptibility to pathogens. Nat. Cell Biol. *15*, 1473–1485.

de Renzis, S., Sönnichsen, B., and Zerial, M. (2002). Divalent Rab effectors regulate the sub-compartmental organization and sorting of early endosomes. Nat. Cell Biol. *4*, 124–133.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology *11*, R25.

Rocha, N., Kuijl, C., van der Kant, R., Janssen, L., Houben, D., Janssen, H., Zwart, W., and Neefjes, J. (2009). Cholesterol sensor ORP1L contacts the ER protein VAP to control Rab7-RILP-p150 Glued and late endosome positioning. J. Cell Biol. *185*, 1209–1225.

Rojas, R., Kametaka, S., Haft, C.R., and Bonifacino, J.S. (2007). Interchangeable but essential functions of SNX1 and SNX2 in the association of retromer with

endosomes and the trafficking of mannose 6-phosphate receptors. Mol. Cell. Biol. *27*, 1112–1124.

Ross, J.L., Ali, M.Y., and Warshaw, D.M. (2008). Cargo transport: molecular motors navigate a complex cytoskeleton. Curr. Opin. Cell Biol. *20*, 41–47.

Rossi, G., Kolstad, K., Stone, S., Palluault, F., and Ferro-Novick, S. (1995). BET3 encodes a novel hydrophilic protein that acts in conjunction with yeast SNAREs. Mol. Biol. Cell *6*, 1769–1780.

Roth, M.G. (2004). Phosphoinositides in constitutive membrane traffic. Physiol. Rev. *84*, 699–730.

Rothman, J.E. (1994). Mechanisms of intracellular protein transport. Nature *372*, 55–63.

Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics *20*, 53–65.

Ruan, J., Dean, A.K., and Zhang, W. (2010). A general co-expression network-based approach to gene expression analysis: comparison and applications. BMC Syst Biol *4*, 8.

Ruprecht, C., Proost, S., Hernandez-Coronado, M., Ortiz-Ramirez, C., Lang, D., Rensing, S.A., Becker, J.D., Vandepoele, K., and Mutwil, M. (2017). Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules. Plant J. *90*, 447–465.

Sacher, M., Barrowman, J., Schieltz, D., Yates, J.R., and Ferro-Novick, S. (2000). Identification and characterization of five new subunits of TRAPP. Eur. J. Cell Biol. *79*, 71–80.

Sacher, M., Kim, Y.-G., Lavie, A., Oh, B.-H., and Segev, N. (2008). The TRAPP complex: insights into its architecture and function. Traffic *9*, 2032–2042.

Saito, A., Hino, S., Murakami, T., Kanemoto, S., Kondo, S., Saitoh, M., Nishimura, R., Yoneda, T., Furuichi, T., Ikegawa, S., et al. (2009). Regulation of endoplasmic reticulum stress response by a BBF2H7-mediated Sec23a pathway is essential for chondrogenesis. Nat. Cell Biol. *11*, 1197–1204.

Sarmah, S., Barrallo-Gimeno, A., Melville, D.B., Topczewski, J., Solnica-Krezel, L., and Knapik, E.W. (2010). Sec24D-dependent transport of extracellular matrix proteins is required for zebrafish skeletal morphogenesis. PLoS ONE *5*, e10367.

Schonteich, E., Wilson, G.M., Burden, J., Hopkins, C.R., Anderson, K., Goldenring, J.R., and Prekeris, R. (2008). The Rip11/Rab11-FIP5 and kinesin II complex regulates endocytic protein recycling. J. Cell. Sci. *121*, 3824–3833.

Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H.D., Schwappach, B., and Weissman, J.S. (2008). The GET complex mediates insertion of

tail-anchored proteins into the ER membrane. Cell *134*, 634–645.

Seaman, M.N.J. (2004). Cargo-selective endosomal sorting for retrieval to the Golgi requires retromer. J. Cell Biol. *165*, 111–122.

Shao, S., and Hegde, R.S. (2011). Membrane protein insertion at the endoplasmic reticulum. Annu. Rev. Cell Dev. Biol. *27*, 25–56.

Sheff, D.R., Daro, E.A., Hull, M., and Mellman, I. (1999). The receptor recycling pathway contains two distinct populations of early endosomes with different sorting functions. J. Cell Biol. *145*, 123–139.

Shin, J.S., and Abraham, S.N. (2001). Co-option of endocytic functions of cellular caveolae by pathogens. Immunology *102*, 2–7.

Shirkhorshidi, A.S., Aghabozorgi, S., and Wah, T.Y. (2015). A Comparison Study on Similarity and Dissimilarity Measures in Clustering Continuous Data. PLOS ONE *10*, e0144059.

Simpson, J.C., Joggerst, B., Laketa, V., Verissimo, F., Cetin, C., Erfle, H., Bexiga, M.G., Singan, V.R., Hériché, J.-K., Neumann, B., et al. (2012). Genome-wide RNAi screening identifies human proteins with a regulatory function in the early secretory pathway. Nat. Cell Biol. *14*, 764–774.

van der Sluijs, P., Hull, M., Webster, P., Mâle, P., Goud, B., and Mellman, I. (1992). The small GTP-binding protein rab4 controls an early sorting event on the endocytic pathway. Cell *70*, 729–740.

Söllner, T., Bennett, M.K., Whiteheart, S.W., Scheller, R.H., and Rothman, J.E. (1993). A protein assembly-disassembly pathway in vitro that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion. Cell *75*, 409–418.

Sönnichsen, B., De Renzis, S., Nielsen, E., Rietdorf, J., and Zerial, M. (2000). Distinct membrane domains on endosomes in the recycling pathway visualized by multicolor imaging of Rab4, Rab5, and Rab11. J. Cell Biol. *149*, 901–914.

Stagg, S.M., Gürkan, C., Fowler, D.M., LaPointe, P., Foss, T.R., Potter, C.S., Carragher, B., and Balch, W.E. (2006). Structure of the Sec13/31 COPII coat cage. Nature *439*, 234–238.

Stanley, P., Taniguchi, N., and Aebi, M. (2015). N-Glycans. In Essentials of Glycobiology, A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, A.G. Darvill, T. Kinoshita, N.H. Packer, J.H. Prestegard, et al., eds. (Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press), p.

Stornaiuolo, M., Lotti, L.V., Borgese, N., Torrisi, M.-R., Mottola, G., Martire, G., and Bonatti, S. (2003). KDEL and KKXX Retrieval Signals Appended to the Same Reporter Protein Determine Different Trafficking between Endoplasmic Reticulum, Intermediate Compartment, and Golgi Complex. Mol Biol Cell *14*, 889–902.

Strahl-Bolsinger, S., Gentzsch, M., and Tanner, W. (1999). Protein O-mannosylation. Biochimica et Biophysica Acta (BBA) - General Subjects *1426*, 297–307.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U.S.A. *102*, 15545–15550.

Südhof, T.C. (2002). Synaptotagmins: Why So Many? J. Biol. Chem. *277*, 7629–7632.

Südhof, T.C., Czernik, A.J., Kao, H.T., Takei, K., Johnston, P.A., Horiuchi, A., Kanazir, S.D., Wagner, M.A., Perin, M.S., and De Camilli, P. (1989). Synapsins: mosaics of shared and individual domains in a family of synaptic vesicle phosphoproteins. Science *245*, 1474–1480.

Sutton, R.B., Fasshauer, D., Jahn, R., and Brunger, A.T. (1998). Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. Nature *395*, 347–353.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447-452.

Sztul, E., and Lupashin, V. (2009). Role of vesicle tethering factors in the ER-Golgi membrane traffic. FEBS Lett. *583*, 3770–3783.

Szul, T., and Sztul, E. (2011). COPII and COPI traffic at the ER-Golgi interface. Physiology (Bethesda) *26*, 348–364.

Tang, L.L., and Balakrishnan, N. (2011). A random-sum Wilcoxon statistic and its application to analysis of ROC and LROC data. J Stat Plan Inference *141*, 335–344.

Taub, F., E., DeLEO, J.M., and Thompson, E.B. (1983). Sequential Comparative Hybridizations Analyzed by Computerized Image Processing Can Identify and Quantitate Regulated RNAs. DNA *2*, 309–327.

Tesson, B.M., Breitling, R., and Jansen, R.C. (2010). DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. BMC Bioinformatics *11*, 497.

Tomoishi, S., Fukushima, S., Shinohara, K., Katada, T., and Saito, K. (2017). CREB3L2-mediated expression of Sec23A/Sec24D is involved in hepatic stellate cell activation through ER-Golgi transport. Sci Rep *7*, 7992.

Traub, L.M. (2009). Tickets to ride: selecting cargo for clathrin-regulated internalization. Nat. Rev. Mol. Cell Biol. *10*, 583–596.

Trimble, W.S., Cowan, D.M., and Scheller, R.H. (1988). VAMP-1: a synaptic vesicle-associated integral membrane protein. Proc. Natl. Acad. Sci. U.S.A. *85*, 4538–4542.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. Science *270*, 484–487.

Vicinanza, M., D'Angelo, G., Di Campli, A., and De Matteis, M.A. (2008). Function and dysfunction of the PI system in membrane trafficking. EMBO J. *27*, 2457–2470.

Vitale, G., Rybin, V., Christoforidis, S., Thornqvist, P., McCaffrey, M., Stenmark, H., and Zerial, M. (1998). Distinct Rab-binding domains mediate the interaction of Rabaptin-5 with GTP-bound Rab4 and Rab5. EMBO J. *17*, 1941–1951.

Vogel, K., and Roche, P.A. (1999). SNAP-23 and SNAP-25 are palmitoylated in vivo. Biochem. Biophys. Res. Commun. *258*, 407–410.

Walseng, E., Bakke, O., and Roche, P.A. (2008). Major histocompatibility complex class II-peptide complexes internalize using a clathrin- and dynamin-independent endocytosis pathway. J. Biol. Chem. *283*, 14717–14727.

Wang, S., Choi, U.B., Gong, J., Yang, X., Li, Y., Wang, A.L., Yang, X., Brunger, A.T., and Ma, C. (2017). Conformational change of syntaxin linker region induced by Munc13s initiates SNARE complex formation in synaptic exocytosis. EMBO J. *36*, 816–829.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet *10*, 57–63.

Weber, T., Zemelman, B.V., McNew, J.A., Westermann, B., Gmachl, M., Parlati, F., Söllner, T.H., and Rothman, J.E. (1998). SNAREpins: minimal machinery for membrane fusion. Cell *92*, 759–772.

van Weering, J.R.T., Verkade, P., and Cullen, P.J. (2010). SNX-BAR proteins in phosphoinositide-mediated, tubular-based endosomal sorting. Semin. Cell Dev. Biol. *21*, 371–380.

Wilcoxon, F. (1946). Individual comparisons of grouped data by ranking methods. J. Econ. Entomol. *39*, 269.

Wolfe, C.J., Kohane, I.S., and Butte, A.J. (2005). Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. BMC Bioinformatics *6*, 227.

Wu, S.-Z., and Bezanilla, M. (2014). Myosin VIII associates with microtubule ends and together with actin plays a role in guiding plant cell division. Elife *3*.

Xu, H., Jun, Y., Thompson, J., Yates, J., and Wickner, W. (2010). HOPS prevents the disassembly of trans-SNARE complexes by Sec17p/Sec18p during membrane fusion. EMBO J. *29*, 1948–1960.

YaDeau, J.T., Klein, C., and Blobel, G. (1991). Yeast signal peptidase contains a glycoprotein and the Sec11 gene product. Proc. Natl. Acad. Sci. U.S.A. *88*, 517–521.

Yamaguchi, T., Dulubova, I., Min, S.-W., Chen, X., Rizo, J., and Südhof, T.C. (2002).

Sly1 binds to Golgi and ER syntaxins via a conserved N-terminal peptide motif. Dev. Cell *2*, 295–305.

Yamaji, T., and Hanada, K. (2015). Sphingolipid metabolism and interorganellar transport: localization of sphingolipid enzymes and lipid transfer proteins. Traffic *16*, 101–122.

Zerial, M., and McBride, H. (2001). Rab proteins as membrane organizers. Nat. Rev. Mol. Cell Biol. *2*, 107–117.