HHS Public Access

Author manuscript

Ann N Y Acad Sci. Author manuscript; available in PMC 2016 September 01.

Published in final edited form as:

Ann N Y Acad Sci. 2015 September ; 1349(1): 83–95. doi:10.1111/nyas.12776.

Human cell adhesion molecules: annotated functional subtypes and overrepresentation of addiction-associated genes

Xiaoming Zhong[1], **Jana Drgonova**[2], **Chuan-Yun Li**[1], and **George R. Uhl**[2,*]

[1]Laboratory of Bioinformatics and Genomic Medicine, Institute of Molecular Medicine, Peking University, Beijing, China 100871

[2]Molecular Neurobiology, NIH-IRP (NIDA), Suite 3510, 333 Cassell Drive Baltimore, Maryland 21224

Abstract

Human cell adhesion molecules (CAMs) are essential both for a) proper development, modulation and maintenance of interactions between cells and for b) cell-to-cell (and matrix-to-cell) communication about these interactions. CAMs are thus key to proper development and plasticity of organs and tissues that include the brain. Despite recognition of the existence of these dual CAM roles and appreciation of the differential functional significance of these roles, there have been surprisingly few systematic studies that have carefully enumerated the universe of CAMs, identified the preferred roles for specific CAMs in distinct types of cellular connections and communication, or related these issues to specific brain disorders or brain circuits. In this paper, we substantially update and review the set of human genes that are likely to encode CAMs based on searches of databases, literature reviews and annotations. We describe the likely CAMs and the functional CAM subclasses into which they fall. These include “iCAMs”, whose contacts largely mediate cell to cell communication, those involved in focal adhesions, CAM genes whose products are preferentially involved with stereotyped and morphologically-identifiable connections between cells (adherens junctions, gap junctions) and smaller numbers of genes in other classes. We discuss a novel proposed mechanism involving selective anchoring of the constituents of iCAM-containing lipid rafts in zones of close neuronal apposition to membranes expressing binding partners of these iCAMs. CAM data from genetic and genomic studies of addiction in humans and mouse models provide examples of the ways in which CAM variation is likely to contribute to a specific brain-based disorder. We discuss how differences in CAM splicing mediated by differences in the addiction-associated splicing regulator RBFOX1/A2BP1 could enrich this picture. CAM expression in dopamine neurons provides one of the ways in which variations in cell adhesion molecule genes could impact a specific set of circuits central to addiction and drug reward.

Keywords

cell adhesion molecules; addiction; dopamine; substance use disorders; lipid rafts; GWAS; connectome

[*]To whom correspondence should be addressed at: Molecular Neurobiology, Box 5180, Baltimore, MD 21224 phone: (410) 550-2843 x 146, fax: (410) 550-1535, guhl@intra.nida.nih.gov.

Introduction

"Cell adhesion molecules" (CAMs) play central roles in much of the connection and communication between cells and their synapses (1–6). Cell adhesion-related communication is essential for many aspects of the proper development of a variety of organs and tissues. This cellular communication also plays substantial roles in the plasticity of cell recognition processes in developed, adult organisms.

Cell adhesion molecules are likely to be especially important in the brain. Proper brain development requires appropriate connection of perhaps 100 trillion synapses (7). Brain function requires substantial plasticity in many of these synapses, providing the bases for learning, memory, addiction and related phenotypes (8,9). Physiologic and cell biologic studies implicate CAM roles in properties that include synapse adhesion (10,11), neuronal connectivity and communication (11), signal transduction (10,12–14), and proper arrangement of pre-synaptic active zones and postsynaptic densities at classical synapses (15,16). We and others have advanced working hypotheses concerning the large contributions of cell adhesion molecules to the development and plasticities of the brain connectome (17), and the CAM "bar codes" that allow the proper connections of specific cell types (GRU and JD, *in preparation*).

Current genetic studies have linked and/or associated variants in cell adhesion molecule genes with a number of phenotypes based on variation in the brain and other organs. Vulnerabilities to addictions are associated with variants in CAM genes in studies of several independent samples (17–21). The importance of CAMs in learning and memory-associated disorders is also demonstrated in genome wide association (GWAS) data (22,23). Genetic variants of CAM genes have been associated with autism (9,24,25). Variants in neuregulin have been associated with vulnerability to schizophrenia (26,27). Variants in a CAM KIAA0319 have been associated with dyslexia (28–30).

Despite the importance of cell adhesion molecules in the normal physiologies of and in the disorders of brain and other organs, and our initial work in defining a set of these genes (2), there remains only a modest amount of updated, systematic work that: 1) enumerates the genes and gene families that function as CAMs; 2) delineates those more likely to function in proper development, modulation and maintenance of morphologically-visible sites for physical interactions between cells and between cells and matrix *vs* those "iCAMs" that appear to largely transmit information about cell-cell and cell-matrix interactions; 3) establishes the ways in which the patterns of CAM expression by any specific cell type might relate to these cells' connectivities and functions; 4) documents the ways in which CAM variation, taken as a whole, might relate to individual differences in vulnerabilities to disease and 5) explores ways in which CAM expression by specific cell types might relate to disease vulnerabilities.

We now report compilation of an updated list of potential human genes annotated or otherwise identified as possible CAMs. We annotate the members of this list that are likely to be CAMs *vs* those that are questionable *vs* those unlikely to be CAMs. For the genes that

are likely to encode *bona fide* CAMs, we describe those likely to play largely information transmission roles between cells ("iCAMs") or between cellular elements and extracellular matrix (*eg* focal adhesions). We contrast these genes to those more likely to be involved in relatively stereotypical, morphologically-visible connections between cells (*eg* adherens junctions, gap junctions). As a specific example of involvement in a complex disorder, we focus on CAMs identified by genome-wide association (GWAS) signals for addiction phenotypes that are both reproducible and modest in individual samples. This list of genes includes many that are expressed in the dopaminergic neurons that play central roles in current models of the reward that can come from abused drugs of many pharmacological classes. These data allow specific hypotheses about the differential connectivities and architectures of dopaminergic neurons in individuals who may display higher *vs* lower expression of (and/or different versions of) interesting cell adhesion molecules. Possible novel roles for glycosylphosphatidyl inositol (GPI)-coupled and other lipid-raft associated CAMs in stabilizing raft contents near areas of close cell-cell apposition are described, providing additional testable hypotheses that flow from our current understanding of the roles for these CAMs. We underscore some of the ways in which understanding CAMs and their human variants is likely to aid understanding of both the brain connectome and a variety of human brain disorders, including addiction.

Identification of human CAM genes

Human CAM gene candidates were identified based on compilation of data from several sources (Fig 1):

1. Entrez Gene query "cell adhesion molecule AND Homo sapiens [organism]".
2. Interpro was searched for genes that encoded common protein domains for CAM families based on common motifs from cadherin, immunoglobulin, fibronectin, integrin, neurexin, neuroligin, cub/sushi and catenin families.
3. The Gene Ontology term "cell adhesion" (GO:0007155) (31) was searched.
4. Our previously-described OKCAM database (2,32) was searched.

We manually curated these candidate CAM gene lists. For each gene, we evaluated evidence from all NCBI data sources that its product(s) were likely to serve as cell adhesion molecule(s), could questionably play such a role, or were unlikely to be cell adhesion molecules. Many of the genes placed in the latter "unlikely" category received "cell adhesion" annotations in other databases due to the gene products' abilities to interact with a cell adhesion molecule, by regulating its expression, for example. We assigned a category based on both the amount and nature of evidence available for each gene (data available at http://rhesusbase.org/OKCAM/).

Data annotations

To elucidate the functions of CAMs, further detailed annotations were assigned to each "likely" CAM gene (Table I). For these genes identified as likely CAMs, we sought evidence that might separate them into functional classes based on a) their involvement in relatively stereotypical and morphologically-recognizable cell-cell contacts, including tight

junctions, gap junctions, desmosomes and adherens junctions; b) their predominant roles in axonal guidance; c) their apparently and/or likely greater roles in transmitting information about cell-cell or cell matrix (focal adhesions) contacts than in mediating physical cell-cell/cell-matrix contacts or their preferential roles in a number of other smaller categories (Table I; Fig 2). We term the products of the genes in the third group "iCAMs" to denote their preferential role in communication as opposed to the establishment of physical interconnections (*but see below*). Online annotations (http://rhesusbase.org/OKCAM/) provide information about expression, regulation, functions, gene structure, genetic variations, phenotype associations, disease associations and drug development for each gene. In Table II, we list the 39 likely CAMs for which SNPs are likely to knock out their expression.

Comparison: human addiction phenotype association GWAS dataset

Data from 500000 – 1M single nucleotide polymorphism (SNP) genome wide association studies for addiction-related phenotypes allowed ranking of genes based on the consistency of their identification by modest genome wide association (GWAS) signals (17). These GWAS signals were provided by clusters of at least 4 SNPs that lay within 10kb of each other and displayed $10^{-2} > p > 10^{-8}$ nominal significance for assessments of case *vs* control allele frequency differences. Nine hundred seventy nine genes contained clusters of SNPs that displayed such nominally-significant case control differences in at least three independent samples were identified. In classical genetic (*eg* twin) studies, the addiction-related phenotypes examined in these studies display substantial evidence for genetic overlap (17). The fraction of the genome (within genes) identified in this way in each independent study provides a basis for identifying the extent to which multiple independent samples would identify the gene by chance. GWAS samples available for ranking these CAM genes include data from eight samples for dependence described in (21,33–37) and five samples studying individual differences in ability to quit smoking (which displays strongly overlapping genetic influences in twin data) in (35,38–42).

RESULTS

1138 candidate CAM genes were identified by one or more of the approaches used here (Fig 1). The Entrez Gene query "cell adhesion molecule AND Homo sapiens" identified 819 gene records. Interpro searches for genes that encoded common CAM protein motifs from the cadherin, immunoglobulin, fibronectin, integrin, neurexin, neuroligin, cub/sushi, and catenin families identified 1716 human proteins, which mapped to 418 human genes. The Gene Ontology term "cell adhesion" (GO:0007155) identified 595 gene records. For comparison, our previously-described OKCAM database (2) identified 424 gene records. There were thus 1138 candidate human CAM genes available for annotation.

Annotation of the records of each of these 1138 candidate CAM genes and relevant literature revealed 474 of these genes that were judged likely to encode *bona fide* cell adhesion molecules (Table I). Three hundred forty four were judged to be unlikely to encode cell adhesion molecules, and 320 were questionable (http://rhesusbase.org/OKCAM/; Fig 1). On average, the genes judged to be "likely" cell adhesion molecules were identified by

almost 2 of the four current annotation methods, while other genes were supported by less evidence. Genes classed as “unlikely” to encode cell adhesion molecules often encoded enzymes or transcription factors whose annotations appeared to arise due to their interactions with cell adhesion molecules. Many “questionable” genes were identified only by electronic annotations that provided insufficient data to provide even moderate confidence in *bona fide* roles in cell adhesion processes or in the absence of such roles. Products of other “questionable” genes displayed ambiguous functions. Such ambiguity is prominent for the large family of collagen genes, whose products contribute to the extracellular matrix which, usually when studded with more specific cell adhesion molecules, can play roles in cell/matrix interactions.

We class 283 of the members of the set of 474 “likely” CAM genes as iCAMs. This subset provided the largest subgroup of the likely cell adhesion molecules. iCAMs were judged to be more involved in providing information about the cell’s environment than in participating in a stereotyped contact or axonal guidance. These iCAMs contain protein motifs from a number of classes. We placed most cadherins in this class after some internal debate, though we acknowledge that several cadherins also participate in stereotypical cell adhesions (*see below*).

The remaining 191 genes are distributed into several groups. Eighty six genes’ products are likely to be involved with interactions between cells and the adjacent extracellular matrix; many of these are secreted and likely to be available in the extracellular space. Thirty six genes are involved in tight junctions. Products of 22 genes are identified primarily with cell-cell recognition for *eg* immune cells, though they are likely to play other roles as well. Eighteen genes, largely expressed on cell surfaces, are involved with focal adhesions. Products of 16 genes are so identified with axonal guidance that they are categorized in this way. The products of “axonal guidance” genes are likely to play other roles. Similarly, products of other cell adhesion molecule genes not annotated in this way are also likely to play roles in axonal guidance. Six are involved selectively in adherens junctions.

We can seek patterns whereby the likely cell adhesion molecule genes are identified in genome wide association data for specific brain disorders. These patterns are likely to provide insights into the chemical coding of connectivities by cell types involved in the circuitries that underlie these disorders. We provide an example for dopaminergic neuronal expression from semiquantitative Allen Brain Atlas data for mice. These data are interpreted in the context of our working hypothesis: for specific types of neurons, the patterns of cell adhesion molecule expression provide principal determinants of the ways in which a neuron’s processes contact other cells and contact itself. These patterns of cell adhesion molecule expression also regulate ways in which a neuron’s processes are contacted by processes of other neurons. These genes’ products thus provide the basic building blocks, or “bar code” (*GRU and JD, in preparation*) for the specificity of the brain connectome. These genes’ variants provide basic underpinnings for the circuitry differences that contribute to brain disorders. In disorders in which brains do not display striking gross neuropathological abnormalities, we anticipate that the variants in these genes contribute prominently to individual differences in vulnerability.

Several specifics help to make these points:

Example: representation of CAMs of different classes among genes identified by modest GWAS signals in multiple addiction phenotype case *vs* control series

Substantial genetic contributions to both dependence and ability to quit smoking have strong support from classical genetic approaches that include twin studies (17). However, genome wide association studies for dependence on illegal or legal addictive substances provide few consistent signals that reach $p < 10^{-8}$ Bonferroni-corrected levels of statistical significance (21). Similar conclusions come from studies of individual differences in the abilities to quit smoking, another addiction-related phenotype (39).

In one approach to the conundrum that this GWAS data raises, phenotypes that include the number of cigarettes smoked/day have been studied, identifying acetylcholine receptor and nicotine metabolizing gene variants that reach or approach these high levels of statistical significance in the large samples that are available for this relatively simple phenotype (43–45).

We have also studied genes that are identified in multiple independent addiction case *vs* control samples by clusters of nearby SNPs that display $10^{-2} > p > 10^{-8}$ levels of statistical significance (5,7,17). We have assembled the lists of genes identified by at least 3–4 such SNPs that lie within 10kb of each other in eight studies of dependence and five studies of ability to quit smoking for which we have complete data (Uhl *et al, unpublished observations* and (17)). Nine hundred seventy nine genes are identified by at least three independent samples in this way. One hundred forty six of these genes are identified in at least 6 of these studies and 16 genes are identified in at least 9 of these independent studies.

One of the initially-unanticipated results of these GWAS datasets has been the overrepresentation of cell adhesion molecules, as we have reported using earlier compilations of the lists of these genes (17). We now find 83 genes on each of two lists: 1) the “likely” cell adhesion molecules that we annotate here and 2) lists of genes identified by at least three GWAS case-control comparisons for addiction phenotypes, dependence or ability to quit smoking (Table III).

Based on the 979/20474 fraction of all genes identified in these GWAS datasets (0.0478) and the 474/20474 fraction of all genes identified in the likely cell adhesion molecule dataset (0.02315), we would expect that 22 genes would be identified in both ways by chance. We actually identify 83. Likely cell adhesion molecules, taken as a group, are thus substantially overrepresented among the genes identified by this approach to analyses of GWAS data ($p= 3.17 \times 10^{-25}$, hypergeometric test). What CAM subclasses do these 83 genes fall into? Most of the 83 cell adhesion molecules that are implicated in addiction phenotypes in this fashion are annotated as “iCAMs” (53 of 83; 63% of the total) or involved in focal adhesion/extracellular matrix interactions (20 of 83 or another 25% of the total).

Individual differences in brains predispose to addiction and altered likelihood of success in quitting smoking. One of the most significant overall contributions to such individual

differences appears to come from variation in interactions between cells, likely largely neurons, that derive from individual differences in products of genes that encode CAMs. CAMs that play especial roles in these differences are largely those whose products do not typically form stereotypic, morphologically-identifiable connections. Quantitative differences in CAM "bar codes" (*GRU and JD, in preparation*) thus alter human addiction vulnerabilities.

Example: overall fit between CAMs identified by modest GWAS signals in multiple addiction phenotype case *vs* control series and those expressed by dopaminergic neurons

Semiquantitative information from Alan Brain Atlas *in situ* hybridization images (46) allows us to confirm dopaminergic expression for several of the addiction- associated genes that encode "likely" CAMs. For PTPRD, CLSTN2, CNTNAP2, ASTN1, CNTNAP5, CHL1, CNTN4, CNTN5, CTNND2, EPHB1 and NRXN3, there is relatively high levels of expression in neurons that are highly likely, based on locations and appearance, to be largely dopaminergic. For DAB1, CTNNA2, PTPRM, CTNNA3, DSCAM, NRG1, OPCML and NLGN1, there is substantial but more modest neuronal expression in these Allen Brain Atlas *in situ* hybridization images.

Example: candidate dopaminergic connection differences arising from variations in genes that encode dopaminergically-expressed cell adhesion molecules identified frequently in addiction GWAS

There is increasing information about the influences of common human haplotypes on several of the likely cell adhesion molecule genes that are both identified in at least 6 addiction case-control GWAS datasets by signals of at least modest magnitude and expressed by dopaminergic neurons at moderate or high levels (Table II). For CDH13 and PTPRD, we have identified 60 – 80% individual differences in expression in postmortem human brains that are associated with common 5′ haplotypes in these genes, as well as haplotypes associated with smaller *ca.* 20% differences in CSMD1 expression (JD, GRU *et al, in preparation*). For NRXN3, we and others have identified common mid-to-3′ haplotypes that alter patterns of splicing of key exons, levels of expression and physiologies of NRXN3-expressing circuits (47,48).

One way to focus on the influences of quantitative and/or qualitative differences in expression of these genes is to focus on connections from and to dopaminergic neurons.

Dopaminergic efferents: connections to striatal/accumbens cholinergic interneurons and cortical neurons—CDH13 and PTPRD mRNAs are both expressed in large, presumable cholinergic striatal neurons and in subsets of deeper cerebral cortical neurons that are most abundant in infralimbic, cingulate and entorhinal cortices in mouse (49). Current *in vitro* data supports homophilic CDH13-CDH13 interactions that inhibit process outgrowth and homophilic PTPRD-PTPRD interactions that foster richer process outgrowth (50,51). Dopaminergic connections with subsets of CDH13-expressing ventral striatal neurons, subsets of CDH13-expressing cortical neurons and perhaps the striatal terminals of these subsets of cortical neurons could all be different in individuals with

differences in levels of expression of CDH13 and PTPRD. Interestingly, mice with deletion of CDH13 do display selective cerebral cortical differences in dopamine, its metabolites, and ratios between dopamine and its metabolites (JD, GRU *et al, in preparation*). These results are consistent with the idea that cortical dopaminergic projections are differentially wired in the absence of CDH13.

Dopaminergic afferents: connections from glutamatergic neurons—Several of the group of addiction-associated dopaminergic cell adhesion molecules defined above appear to be expressed postsynaptically by neurons that receive glutamatergic afferents. These include the products of the NLGN1 and CLSTN2 genes as well as the LRRTM3 gene product of the combined CTNNA3_ LRRTM3 locus (52–54). Double labeling experiments have identified glutamatergic synapses on dopamine neurons that come from VTA afferents of neurons whose cell bodies lie in a number of regions (55). These include regions of the prefrontal cortex, lateral and medial hypothalamic and preoptic areas, ventral pallidum, lateral habenula, dorsal and median raphe, mesopontine central gray and reticular formation, pedunculopontine and laterodorsal tegmental nuclei, parabrachial nucleus, cuneiform nucleus and medial septum/diagonal band of Broca. The LRRTM3 and NLGN1 gene products expressed by dopaminergic neurons are thus themselves likely to interact with neurexins and neurexin homologs that include CNTNAPs that are expressed in neurons in several of these zones of origin of glutamatergic VTA afferents. mRNAs for NRXN3 and CNTNAP2 are expressed robustly in neurons in most of these regions, though there are only low levels of CNTNAP5 expression (56). Conceivably, some of the influences of addiction-associated variation in genes encoding neurexin family proteins could come from variation in their expression by glutamatergic VTA afferents that arise from cell bodies in these areas.

Example: candidate addiction-associated splicing differences that could be mediated by RBFOX1 allelic variants

The RBFOX1 (or A2BP1) gene product serves as a regulator of splicing of the primary RNA transcripts of genes whose sequences include its canonical recognition motif (U)GCAUG and at sites that do not display this *cis*-acting element (57). We have identified modest associations of RBFOX1 variants with substance dependence or ability to quit smoking in 13 independent datasets (17,21,33–35,37–39,42,58,59). RBFOX1 variation has also been associated with individual differences in smoking quantity/frequency (44). Genomic markers within or quite near RBFOX1 have displayed linkage to substance dependence related phenotypes (60–62). Mice with RBFOX1 expression deleted from neurons display altered expression of splice variants from twenty genes. These genes include the NRCAM and NRXN3 iCAM genes that display association with substance dependence in several samples (47,63), as well as the iCAM PTPRO.

Shen and colleagues have recently used a variety of approaches to identify nominal high significance for RBFOX1 binding overlaps with epigenetic marks in a set of genes that includes many CAMs (64). These workers identify immunoprecipitation of RBFOX1 with antibodies that recognize the H3K4me3 modified histone (64). These workers and our laboratory have identified changes in cocaine reward with local and brain-wide knockdown/knockout of A2BP1 activities (64) (JD, GRU *et al, in preparation*).

Taken together, these genetic, epigenetic and behavioral results support the working ideas that splicing differences mediated by RBFOX1 allelic variants, likely to include variation in expression of CAMs, contribute to vulnerabilities to addiction.

DISCUSSION

The present review provides substantial updates to the list of human “cell adhesion” molecules, identifying likely, questionable and unlikely candidates from among the longer lists of candidate CAM genes. It seeks, for the first time, to assign cell adhesion molecule gene products to categories that signify more relevance to their adhesive properties *vs* those that appear more relevant to information-transmitting properties that many of these gene products mediate. We apply each of these labels based on reviews of currently published literature and annotations for each gene. These annotations substantially update our prior work with this gene set. As information about many of the “questionable” cell adhesion molecules increases, additional genes may well be recognized as “likely” cell adhesion molecules. As more information about the properties of more of these cell adhesion molecules becomes available, the assignment of the encoded products to “more adhesion” *vs* “more information” will also change. We welcome readers’ comments and will use these comments to update this list and its annotations.

The annual number of PubMed citations for “cell adhesion molecules” grew more than 6 fold between the late 1980s and 1990’s, but has grown by less than 1/3 since then. Cell adhesion molecules are likely to play roles in development and adult function of virtually every cell and tissue. Brain expression of many of these molecules appears to have especial relevance for some of the most challenging problems in normal and pathological biologies: How are the *ca.*100 trillion neuronal connections in the brain established appropriately? How are these connections modified by exposures to different patterns of activity that result from experiences and from exposure to drugs that act on the nervous system? How do differences in these connections and their modification with experience result in brain disorders? This review posits central roles for cell adhesion molecules in the answers to each of these questions. While the overall complexity of the influences of CAMs on the connectome is great, this complexity may be more manageable as we focus on specific cell types, specific circuits that involve these cell types and specific disease processes. Here, we thus focus on the subsets of cell adhesion molecule genes that are identified by modest GWAS signals in multiple independent studies of addiction-related phenotypes. This approach has limits. The lists of genes identified by modest signals in multiple addiction-related GWAS studies are likely to contain both false positives and false negatives in identifying CAM genes whose variations alter individuals’ vulnerabilities to express addiction-related phenotypes. The great diversity of genes identified here is often accompanied by great diversity of splicing variants (48,65), which can both play large roles in information provided by the CAMs expression. Splicing variation introduces complexity that we have tackled from the perspective of RBFOX1, but we are likely to have omitted discussion of other important splicing events. The focus on CAMs in the current review should not obscure roles for other mechanisms that have been postulated to aid in the long-term storage of information about prior drug experiences that provides the key feature of addiction, including changes in cyclic nucleotides (66), G and RGS proteins (67,68),

neurotropic factors (69), patterns of protein phosphorylation (70), transcription factor expression (71), and histone modifications (72).

The set of genes identified here by both dopaminergic expression and addiction phenotype GWAS in multiple samples does provide evidence for biological plausibility. The plausibility of identification of specific sets of cell adhesion molecules, such as those associated with glutamatergic synapse formation, is enhanced by the relatively recent general recognition of the role of glutamate as cotransmitter for subsets of ventral midbrain dopaminergic neurons and the mapping of substantial numbers of glutamatergic neurons among the afferents to these dopaminergic neurons (73). These anatomical relationships are fit by the patterns of expression of cell adhesion molecules attributed to glutamatoceptive and glutamatergic neurons in the Alan Brain Atlas data for dopaminergic neurons, and for recipients of the efferent connections from dopamine neurons and the sources of many of their afferents.

GPI-anchored cell adhesion molecules are similarly expressed in targets of midbrain dopaminergic efferents and by neurons in regions that provide afferents to midbrain dopamine neurons. Localization of the GPI-anchored (and, likely, other) cell adhesion molecules to lipid raft domains supports an idea that has not been stated clearly previously, to our knowledge: that one of the roles of interactions between cell adhesion molecule binding partners located on adjacent neuronal processes might be to stabilize the lipid rafts that contain them in proximity to each other (Fig 3). Lipid rafts contain not only cell adhesion molecules but also transporters, G-protein coupled receptors, channels and G proteins (74,75). We can think about “presynaptic” and “postsynaptic” lipid raft pairs that are likely to be stabilized when the cell membrane elements that express them are close enough to allow interactions between the GPI-anchored and other cell adhesion molecules contained in each raft in the pair. Proximities that could facilitate such interactions can often be found in perisynaptic regions adjacent to classical synaptic specializations. Biochemical evidence supports localization of biochemically- and morphologically-defined lipid rafts that contain cell adhesion molecules next to the classical synaptic specializations (76). Other sites could also display important CAM recognition/paired raft stabilization. Monoaminergic projections to cortex are characterized by varicosities that contain clusters of apparent synaptic vesicles (77). These varicosities are often found in regions of close apposition of membranes that do not display electron densities after the heavy metal salt stains that characterize classical synapses. These likely nonclassical “synapses” also appear to be strong candidates for stabilization by CAMs expressed in lipid raft pairs from the monoaminergic projections and the cortical neuronal recipients of this innervation. The roles that GPI-anchored, and other cell adhesion molecules could play in stabilizing raft pairs at classical and nonclassical “synapses” imply their indirect roles in stabilizing the other contents of these raft pairs, (*eg* transporters, channels *etc*) in proximity to each other. To the extent that individual differences in CDH13, CNTN4 and CNTN5 expression alter the abundance of such closely-approximated “pre/perisynaptic-post/perisynaptic” lipid raft pairs in neurons that express them, the other contents of these rafts could be assembled in different abundance with significant impacts on dopaminergic functions.

Other GPI-anchored cell adhesion molecules are likely to decorate lipid rafts that are not stabilized due to expression in closely-approximated membrane domains. These cell adhesion molecules could still recognize the soluble fragments that are produced from many cell adhesion molecule genes, including interesting soluble fragment products of a CDH13 splicing variant (JD, GRU *et al, in preparation*).

The cell adhesion molecules identified and categorized here thus have functions beyond just "cell glue" in ways that can generate specific testable hypotheses about their roles in specific cell types and specific circuits, and may even provide substrates for novel therapeutics that can modify brain connections (78). These hypotheses should be assessed in light of the strengths and limitations of the approaches used here, and the strengths and limitations of the underlying datasets employed for these analyses. Cell adhesion molecules and mechanisms remain fascinating and understudied ways in which the body, and the brain in particular, assembles and changes its assemblies during its development and through its interactions with the environment.

Acknowledgments

We acknowledge support by the NIH IRP (NIDA) (GRU), from the National Key Basic Research Program of China [2013CB531202] (CYL), from each of the investigators and subjects of the GWAS studies whose data is summarized here, and helpful comments on the manuscript from D Martinelli.

References

1. Winograd-Katz SE, Fassler R, Geiger B, Legate KR. The integrin adhesome: from genes and proteins to human disease. Nat Rev Mol Cell Biol. 2014; 15:273–288. [PubMed: 24651544]
2. Li CY, Liu QR, Zhang PW, Li XM, Wei L, Uhl GR. OKCAM: an ontology-based, human-centered knowledgebase for cell adhesion molecules. Nucleic Acids Res. 2009; 37:D251–260. [PubMed: 18790807]
3. Clarke RA, Eapen V. Balance within the Neurexin Trans-Synaptic Connexus Stabilizes Behavioral Control. Front Hum Neurosci. 2014; 8:52. [PubMed: 24578685]
4. Baker MW, Macagno ER. Control of neuronal morphology and connectivity: emerging developmental roles for gap junctional proteins. FEBS Lett. 2014; 588:1470–1479. [PubMed: 24560787]
5. Sheng L, Leshchyns'ka I, Sytnyk V. Cell adhesion and intracellular calcium signaling in neurons. Cell Commun Signal. 2013; 11:94. [PubMed: 24330678]
6. Yamada S, Nelson WJ. Synapses: sites of cell recognition, adhesion, and functional specification. Annu Rev Biochem. 2007; 76:267–294. [PubMed: 17506641]
7. Pakkenberg B, Pelvig D, Marner L, Bundgaard MJ, Gundersen HJ, Nyengaard JR, Regeur L. Aging and the human neocortex. Exp Gerontol. 2003; 38:95–99. [PubMed: 12543266]
8. Hishimoto A, Liu QR, Drgon T, Pletnikova O, Walther D, Zhu XG, Troncoso JC, Uhl GR. Neurexin 3 polymorphisms are associated with alcohol dependence and altered expression of specific isoforms. Hum Mol Genet. 2007; 16:2880–2891. [PubMed: 17804423]
9. Kim HG, Kishikawa S, Higgins AW, Seong IS, Donovan DJ, Shen Y, Lally E, Weiss LA, Najm J, Kutsche K, et al. Disruption of neurexin 1 associated with autism spectrum disorder. Am J Hum Genet. 2008; 82:199–207. [PubMed: 18179900]
10. Shapiro L, Love J, Colman DR. Adhesion molecules in the nervous system: structural insights into function and diversity. Annu Rev Neurosci. 2007; 30:451–474. [PubMed: 17600523]
11. Yamada S, Nelson WJ. Synapses: sites of cell recognition, adhesion, and functional specification. Annu Rev Biochem. 2007; 76:267–294. [PubMed: 17506641]

12. Stoker AW. Protein tyrosine phosphatases and signalling. J Endocrinol. 2005; 185:19–33. [PubMed: 15817824]
13. Salinas PC, Price SR. Cadherins and catenins in synapse development. Curr Opin Neurobiol. 2005; 15:73–80. [PubMed: 15721747]
14. Hirano S, Suzuki ST, Redies C. The cadherin superfamily in neural development: diversity, function and interaction with other molecules. Front Biosci. 2003; 8:d306–355. [PubMed: 12456358]
15. Song JY, Ichtchenko K, Sudhof TC, Brose N. Neuroligin 1 is a postsynaptic cell-adhesion molecule of excitatory synapses. Proc Natl Acad Sci U S A. 1999; 96:1100–1105. [PubMed: 9927700]
16. Dityatev A, Dityateva G, Schachner M. Synaptic strength as a function of post- versus presynaptic expression of the neural cell adhesion molecule NCAM. Neuron. 2000; 26:207–217. [PubMed: 10798405]
17. Uhl GR, Drgon T, Johnson C, Li CY, Contoreggi C, Hess J, Naiman D, Liu QR. Molecular genetics of addiction and related heritable phenotypes: genome-wide association approaches identify “connectivity constellation” and drug target genes with pleiotropic effects. Ann N Y Acad Sci. 2008; 1141:318–381. [PubMed: 18991966]
18. Johnson C, Drgon T, Liu QR, Walther D, Edenberg H, Rice J, Foroud T, Uhl GR. Pooled association genome scanning for alcohol dependence using 104,268 SNPs: validation and use to identify alcoholism vulnerability loci in unrelated individuals from the collaborative study on the genetics of alcoholism. Am J Med Genet B Neuropsychiatr Genet. 2006; 141B:844–853. [PubMed: 16894614]
19. Liu QR, Drgon T, Johnson C, Walther D, Hess J, Uhl GR. Addiction molecular genetics: 639,401 SNP whole genome association identifies many “cell adhesion” genes. Am J Med Genet B Neuropsychiatr Genet. 2006; 141:918–925. [PubMed: 17099884]
20. Uhl GR, Liu QR, Drgon T, Johnson C, Walther D, Rose JE, David SP, Niaura R, Lerman C. Molecular genetics of successful smoking cessation: convergent genome-wide association study results. Arch Gen Psychiatry. 2008; 65:683–693. [PubMed: 18519826]
21. Johnson C, Drgon T, Walther D, Uhl GR. Genomic regions identified by overlapping clusters of nominally-positive SNPs from genome-wide studies of alcohol and illegal substance dependence. PLoS One. 2011; 6:e19210. [PubMed: 21818250]
22. Butcher LM, Meaburn E, Dale PS, Sham P, Schalkwyk LC, Craig IW, Plomin R. Association analysis of mild mental impairment using DNA pooling to screen 432 brain-expressed single-nucleotide polymorphisms. Mol Psychiatry. 2005; 10:384–392. [PubMed: 15452586]
23. Hargreaves A, Anney R, O’Dushlaine C, Nicodemus KK, Gill M, Corvin A, Morris D, Donohoe G. The one and the many: effects of the cell adhesion molecule pathway on neuropsychological function in psychosis. Psychol Med. 2013:1–11.
24. Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, Ikeda M, Rea A, Guy M, Lin S, Cook EH, et al. A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. Am J Hum Genet. 2008; 82:160–164. [PubMed: 18179894]
25. Girirajan S, Dennis MY, Baker C, Malig M, Coe BP, Campbell CD, Mark K, Vu TH, Alkan C, Cheng Z, et al. Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. Am J Hum Genet. 2013; 92:221–237. [PubMed: 23375656]
26. Munafo MR, Attwood AS, Flint J. Neuregulin 1 genotype and schizophrenia. Schizophr Bull. 2008; 34:9–12. [PubMed: 18032396]
27. Bernstein HG, Bogerts B. Neuregulin-1 alpha, the underestimated molecule: emerging new roles in normal brain function and the pathophysiology of schizophrenia? Genome. 2013; 56:703–704. [PubMed: 24299109]
28. Velayos-Baeza A, Toma C, da Roza S, Paracchini S, Monaco AP. Alternative splicing in the dyslexia-associated gene KIAA0319. Mamm Genome. 2007; 18:627–634. [PubMed: 17846832]
29. Paracchini S, Thomas A, Castro S, Lai C, Paramasivam M, Wang Y, Keating BJ, Taylor JM, Hacking DF, Scerri T, et al. The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. Hum Mol Genet. 2006; 15:1659–1666. [PubMed: 16600991]

30. Mascheretti S, Riva V, Giorda R, Beri S, Lanzoni LF, Cellino MR, Marino C. KIAA0319 and ROBO1: evidence on association with reading and pleiotropic effects on language and mathematics abilities in developmental dyslexia. J Hum Genet. 2014; 59:189–197. [PubMed: 24430574]
31. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25:25–29. [PubMed: 10802651]
32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25:25–29. [PubMed: 10802651]
33. Uhl GR, Drgon T, Liu QR, Johnson C, Walther D, Komiyama T, Harano M, Sekine Y, Inada T, Ozaki N, et al. Genome-wide association for methamphetamine dependence: convergent results from 2 samples. Arch Gen Psychiatry. 2008; 65:345–355. [PubMed: 18316681]
34. Drgon T, Johnson CA, Nino M, Drgonova J, Walther DM, Uhl GR. “Replicated” genome wide association for dependence on illegal substances: genomic regions identified by overlapping clusters of nominally positive SNPs. American Journal of Medical Genetics Part B Neuropsychiatric Genetics. 2011; 156:125–138.
35. Uhl GR, Walther D, Musci R, Fisher C, Anthony JC, Storr CL, Behm FM, Eaton WW, Ialongo N, Rose JE. Smoking quit success genotype score predicts quit success and distinct patterns of developmental involvement with common addictive substances. Mol Psychiatry. 2014; 19:50–54. [PubMed: 23128154]
36. Johnson C, Drgon T, Liu QR, Zhang PW, Walther D, Li CY, Anthony JC, Ding Y, Eaton WW, Uhl GR. Genome wide association for substance dependence: convergent results from epidemiologic and research volunteer samples. BMC Med Genet. 2008; 9:113. [PubMed: 19094236]
37. Drgon T, Montoya I, Johnson C, Liu QR, Walther D, Hamer D, Uhl GR. Genome-wide association for nicotine dependence and smoking cessation success in NIH research volunteers. Mol Med. 2009; 15:21–27. [PubMed: 19009022]
38. Drgon T, Johnson C, Walther D, Albino AP, Rose JE, Uhl GR. Genome-wide association for smoking cessation success: participants in a trial with adjunctive denicotinized cigarettes. Mol Med. 2009; 15:268–274. [PubMed: 19593411]
39. Uhl GR, Liu QR, Drgon T, Johnson C, Walther D, Rose JE, David SP, Niaura R, Lerman C. Molecular genetics of successful smoking cessation: convergent genome-wide association study results. Arch Gen Psychiatry. 2008; 65:683–693. [PubMed: 18519826]
40. Uhl GR, Walther D, Musci R, Fisher C, Anthony JC, Storr CL, Behm FM, Eaton WW, Ialongo N, Rose JE. Smoking quit success genotype score predicts quit success and distinct patterns of developmental involvement with common addictive substances. Mol Psychiatry. 2012
41. Uhl GR, Drgon T, Johnson C, Ramoni MF, Behm FM, Rose JE. Genome-wide association for smoking cessation success in a trial of precessation nicotine replacement. Mol Med. 2010; 16:513–526. [PubMed: 20811658]
42. Uhl GR, Drgon T, Johnson C, Walther D, David SP, Aveyard P, Murphy M, Johnstone EC, Munafo MR. Genome-wide association for smoking cessation success: participants in the Patch in Practice trial of nicotine replacement. Pharmacogenomics. 2010; 11:357–367. [PubMed: 20235792]
43. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature. 2008; 452:638–642. [PubMed: 18385739]
44. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, Sulem P, Rafnar T, Esko T, Walter S, et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nat Genet. 2010; 42:448–453. [PubMed: 20418888]
45. Bierut LJ. Convergence of genetic findings for nicotine dependence and smoking related diseases with chromosome 15q24-25. Trends Pharmacol Sci. 2010; 31:46–51. [PubMed: 19896728]
46. http://mouse.brain-map.org/

47. Hishimoto A, Liu QR, Drgon T, Pletnikova O, Walther D, Zhu XG, Troncoso JC, Uhl GR. Neurexin 3 polymorphisms are associated with alcohol dependence and altered expression of specific isoforms. Hum Mol Genet. 2007; 16:2880–2891. [PubMed: 17804423]
48. Aoto J, Martinelli DC, Malenka RC, Tabuchi K, Sudhof TC. Presynaptic neurexin-3 alternative splicing trans-synaptically controls postsynaptic AMPA receptor trafficking. Cell. 2013; 154:75–88. [PubMed: 23827676]
49. mouse.brain-map.org
50. Ciatto C, Bahna F, Zampieri N, VanSteenhouse HC, Katsamba PS, Ahlsen G, Harrison OJ, Brasch J, Jin X, Posy S, et al. T-cadherin structures reveal a novel adhesive binding mechanism. Nat Struct Mol Biol. 2010; 17:339–347. [PubMed: 20190755]
51. Wang J, Bixby JL. Receptor tyrosine phosphatase-delta is a homophilic, neurite-promoting cell adhesion molecular for CNS neurons. Mol Cell Neurosci. 1999; 14:370–384. [PubMed: 10588391]
52. Craig AM, Kang Y. Neurexin-neuroligin signaling in synapse development. Curr Opin Neurobiol. 2007; 17:43–52. [PubMed: 17275284]
53. Uvarov P, Kajander T, Airaksinen MS. Origin and loss of nested LRRTM/alpha-catenin genes during vertebrate evolution. PLoS One. 2014; 9:e89910. [PubMed: 24587117]
54. Pettem KL, Yokomaku D, Luo L, Linhoff MW, Prasad T, Connor SA, Siddiqui TJ, Kawabe H, Chen F, Zhang L, et al. The specific alpha-neurexin interactor calsyntenin-3 promotes excitatory and inhibitory synapse development. Neuron. 2013; 80:113–128. [PubMed: 24094106]
55. Geisler S, Derst C, Veh RW, Zahm DS. Glutamatergic afferents of the ventral tegmental area in the rat. J Neurosci. 2007; 27:5730–5743. [PubMed: 17522317]
56. mouse.brain-map.org
57. Gehman LT, Stoilov P, Maguire J, Damianov A, Lin CH, Shiue L, Ares M Jr, Mody I, Black DL. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. Nat Genet. 2011; 43:706–711. [PubMed: 21623373]
58. Liu QR, Drgon T, Johnson C, Walther D, Hess J, Uhl GR. Addiction molecular genetics: 639,401 SNP whole genome association identifies many “cell adhesion” genes. American Journal of Medical Genetics Part B Neuropsychiatric Genetics. 2006; 141B:918–925.
59. Uhl GR, Liu QR, Drgon T, Johnson C, Walther D, Rose JE. Molecular genetics of nicotine dependence and abstinence: whole genome association using 520,000 SNPs. BMC Genet. 2007; 8:10. [PubMed: 17407593]
60. Long JC, Knowler WC, Hanson RL, Robin RW, Urbanek M, Moore E, Bennett PH, Goldman D. Evidence for genetic linkage to alcohol dependence on chromosomes 4 and 11 from an autosome-wide scan in an American Indian population. Am J Med Genet. 1998; 81:216–221. [PubMed: 9603607]
61. Ehlers CL, Gilder DA, Wall TL, Phillips E, Feiler H, Wilhelmsen KC. Genomic screen for loci associated with alcohol dependence in Mission Indians. Am J Med Genet B Neuropsychiatr Genet. 2004; 129B:110–115. [PubMed: 15274051]
62. Morley KI, Medland SE, Ferreira MA, Lynskey MT, Montgomery GW, Heath AC, Madden PA, Martin NG. A possible smoking susceptibility locus on chromosome 11p12: evidence from sex-limitation linkage analyses in a sample of Australian twin families. Behav Genet. 2006; 36:87–99. [PubMed: 16365831]
63. Ishiguro H, Liu QR, Gong JP, Hall FS, Ujike H, Morales M, Sakurai T, Grumet M, Uhl GR. NrCAM in addiction vulnerability: positional cloning, drug-regulation, haplotype-specific expression, and altered drug reward in knockout mice. Neuropsychopharmacology. 2006; 31:572–584. [PubMed: 16123759]
64. Feng J, Wilkinson M, Liu X, Purushothaman I, Ferguson D, Vialou V, Maze I, Shao N, Kennedy P, Koo J, et al. Chronic cocaine-regulated epigenomic changes in mouse nucleus accumbens. Genome Biol. 2014; 15:R65. [PubMed: 24758366]
65. Treutlein B, Gokce O, Quake SR, Sudhof TC. Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. Proc Natl Acad Sci U S A. 2014; 111:E1291–1299. [PubMed: 24639501]

66. Terwilliger RZ, Beitner-Johnson D, Sevarino KA, Crain SM, Nestler EJ. A general role for adaptations in G-proteins and the cyclic AMP system in mediating the chronic actions of morphine and cocaine on neuronal function. Brain Res. 1991; 548:100–110. [PubMed: 1651140]

67. Nestler EJ, Erdos JJ, Terwilliger R, Duman RS, Tallman JF. Regulation of G proteins by chronic morphine in the rat locus coeruleus. Brain Res. 1989; 476:230–239. [PubMed: 2495149]

68. Zachariou V, Georgescu D, Sanchez N, Rahman Z, DiLeone R, Berton O, Neve RL, Sim-Selley LJ, Selley DE, Gold SJ, et al. Essential role for RGS9 in opiate action. Proc Natl Acad Sci U S A. 2003; 100:13656–13661. [PubMed: 14595021]

69. Berhow MT, Russell DS, Terwilliger RZ, Beitner-Johnson D, Self DW, Lindsay RM, Nestler EJ. Influence of neurotrophic factors on morphine- and cocaine-induced biochemical changes in the mesolimbic dopamine system. Neuroscience. 1995; 68:969–979. [PubMed: 8545003]

70. Guitart X, Nestler EJ. Second messenger and protein phosphorylation mechanisms underlying opiate addiction: studies in the rat locus coeruleus. Neurochem Res. 1993; 18:5–13. [PubMed: 8385277]

71. Hope B, Kosofsky B, Hyman SE, Nestler EJ. Regulation of immediate early gene expression and AP-1 binding in the rat nucleus accumbens by chronic cocaine. Proc Natl Acad Sci U S A. 1992; 89:5764–5768. [PubMed: 1631058]

72. Maze I, Covington HE 3rd, Dietz DM, LaPlant Q, Renthal W, Russo SJ, Mechanic M, Mouzon E, Neve RL, Haggarty SJ, et al. Essential role of the histone methyltransferase G9a in cocaine-induced plasticity. Science. 2010; 327:213–216. [PubMed: 20056891]

73. Trudeau LE, Hnasko TS, Wallen-Mackenzie A, Morales M, Rayport S, Sulzer D. The multilingual nature of dopamine neurons. Prog Brain Res. 2014; 211:141–164. [PubMed: 24968779]

74. Yu H, Wakim B, Li M, Halligan B, Tint GS, Patel SB. Quantifying raft proteins in neonatal mouse brain by 'tube-gel' protein digestion label-free shotgun proteomics. Proteome Sci. 2007; 5:17. [PubMed: 17892558]

75. Williamson R, Thompson AJ, Abu M, Hye A, Usardi A, Lynham S, Anderton BH, Hanger DP. Isolation of detergent resistant microdomains from cultured neurons: detergent dependent alterations in protein composition. BMC Neurosci. 2010; 11:120. [PubMed: 20858284]

76. Suzuki T, Zhang J, Miyazawa S, Liu Q, Farzan MR, Yao WD. Association of membrane rafts and postsynaptic density: proteomics, biochemical, and ultrastructural analyses. J Neurochem. 2011; 119:64–77. [PubMed: 21797867]

77. Carr DB, O'Donnell P, Card JP, Sesack SR. Dopamine terminals in the rat prefrontal cortex synapse on pyramidal cells that project to the nucleus accumbens. J Neurosci. 1999; 19:11049–11060. [PubMed: 10594085]

78. Uhl GR, Drgonova J. Cell adhesion molecules: druggable targets for modulating the connectome and brain disorders? Neuropsychopharmacology. 2014; 39:235. [PubMed: 24317312]

79. Yang H, He B, Ma H, Tsaur SC, Ma C, Wu Y, Ting CT, Zhang YE. Expression profile and gene age jointly shaped the genome-wide distribution of premature termination codons in a Drosophila melanogaster population. Mol Biol Evol. 2014

80. Drgon T, Johnson CA, Nino M, Drgonova J, Walther DM, Uhl GR. "Replicated" genome wide association for dependence on illegal substances: genomic regions identified by overlapping clusters of nominally positive SNPs. Am J Med Genet B Neuropsychiatr Genet. 2011; 156:125–138. [PubMed: 21302341]

81. Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P, Maier W, Moessner R, Gaebel W, Dahmen N, et al. Genome-wide association study of alcohol dependence. Arch Gen Psychiatry. 2009; 66:773–784. [PubMed: 19581569]

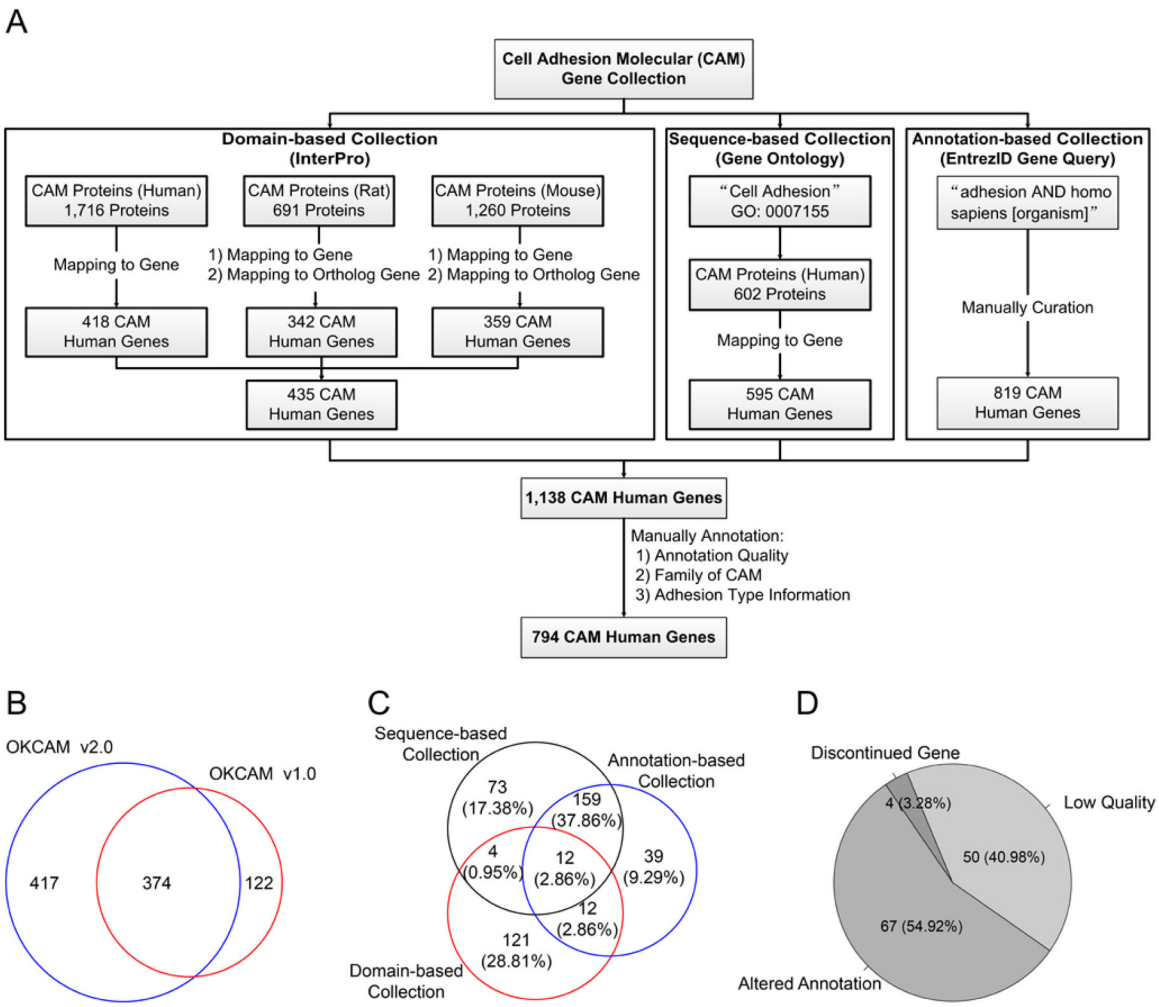

Figure 1. Cell adhesion molecule gene identification and annotation

Identification of human cell adhesion molecules.

(A) Cell adhesion molecules were compiled by integrating Gene Ontology annotations, domain structure information and keyword queries against NCBI Entrez annotations. 794 unique human genes were identified after manual curation. (B) Overlap of the current version of OKCAM with the previous version. (C) Characteristics of the 417 newly-added CAMs. (D) Characteristics of the122 genes that were included in the prior CAM dataset but not included in the current set.

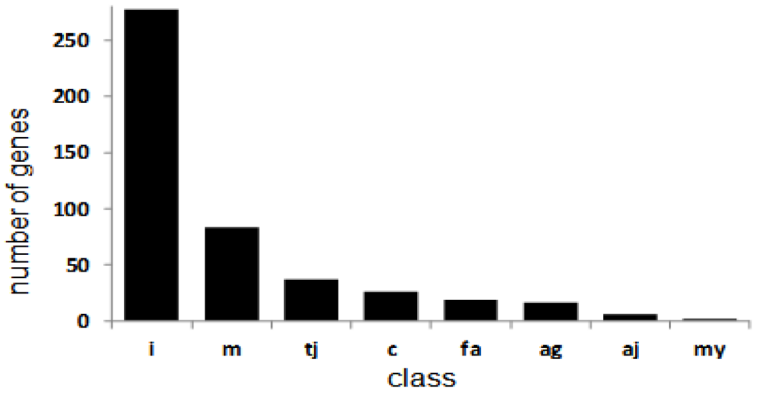

Figure 2. Functional classes of genes encoding likely CAMs

Distribution of likely CAM genes into classes annotated here. CAM types: i: information predominant CAM, m: primarily involved in interactions with cell matrix, tj: primary involvement in tight junctions c: primary roles in cell/cell interactions, principally in immune system, fa: primary involvement in focal adhesions, ag: primary roles in axonal guidance, aj: primary role in adherens junctions, my: primarily involved in myelin interactions. Please note likely involvement of many of the products of these CAM genes in multiple functions (*esp* cadherins).

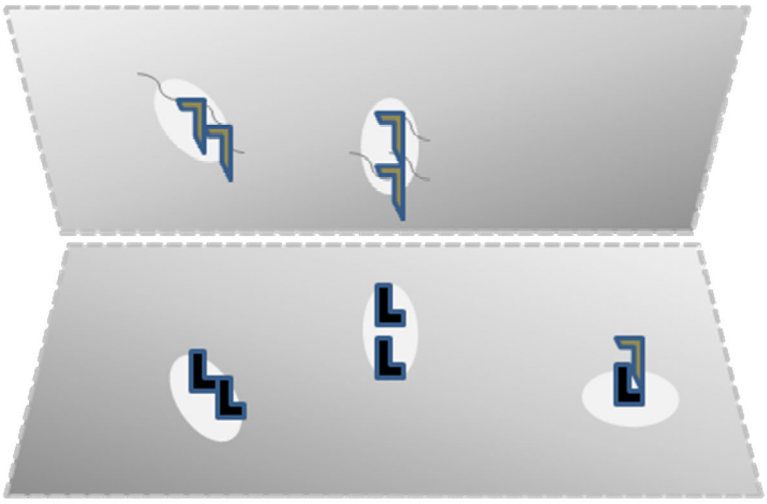

Figure 3.

Schematic of pre (*eg* top) and post (*eg* bottom) membranes pulled away from each other to illustrate the way in which CAMs (L shapes) anchored to lipid rafts (ovals) might stabilize the constituents of the lipid rafts (squiggles) by allowing binding between “pre” and ”postsynaptic” CAM-containing lipid raft pairs. Bottom right: Lipid raft containing CAM bound to soluble CAM, blocking possible participation in stabilization of a lipid raft pair.

Table I

Genes judged “likely” to encode *bona fide* human cell adhesion molecules.

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
CDH1	999	i	1	1	1	3	cadherin 1
CDH10	1008	i	1	1	1	3	cadherin 10
CDH11	1009	i	1	1	1	3	cadherin 11
CDH13	1012	i	1	1	1	3	cadherin 13
CDH15	1013	i	1	1	1	3	cadherin 15
CDH16	1014	i	1	1	1	3	cadherin 16
CDH17	1015	i	1	1	1	3	cadherin 17
CDH2	1000	i	1	1	1	3	cadherin 2
CDH23	64072	i	1	1	1	3	cadherin-related 23
CDH24	64403	i	1	1	1	3	cadherin 24
CDH3	1001	i	1	1	1	3	cadherin 3
CDH4	1002	i	1	1	1	3	cadherin 4
CDH6	1004	i	1	1	1	3	cadherin 6
CDH7	1005	i	1	1	1	3	cadherin 7
CDH8	1006	i	1	1	1	3	cadherin 8
CDH9	1007	i	1	1	1	3	cadherin 9
CDHR2	54825	i	1	1	1	3	cadherin-related family member 2
CDHR5	53841	i	1	1	1	3	cadherin-related family member 5
CEACAM1	634	i	1	1	1	3	carcinoembryonic antigen-related cell adhesion molecule 1
CELSR2	1952	i	1	1	1	3	cadherin EGF LAG seven-pass G-type receptor 2
CLSTN1	22883	i	1	1	1	3	calsyntenin 1
CNTN4	152330	i	1	1	1	3	contactin 4
DCHS1	8642	i	1	1	1	3	dachsous cadherin-related 1
DSCAM	1826	i	1	1	1	3	Down syndrome cell adhesion molecule
DSCAML1	57453	i	1	1	1	3	Down syndrome cell adhesion molecule like 1
EPHA2	1969	i	1	1	1	3	EPH receptor A2
EPHA3	2042	i	1	1	1	3	EPH receptor A3
EPHA7	2045	i	1	1	1	3	EPH receptor A7
EPHA8	2046	i	1	1	1	3	EPH receptor A8

Ann N Y Acad Sci. Author manuscript; available in PMC 2016 September 01.

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
EPHB3	2049	i	1	1	1	3	EPH receptor B3
EPHB4	2050	i	1	1	1	3	EPH receptor B4
FAT1	2195	i	1	1	1	3	FAT atypical cadherin 1
FAT2	2196	i	1	1	1	3	FAT atypical cadherin 2
FAT4	79633	i	1	1	1	3	FAT atypical cadherin 4
L1CAM	3897	i	1	1	1	3	L1 cell adhesion molecule
LRRN2	10446	i	1	1	1	3	leucine rich repeat neuronal 2
NCAM2	4685	i	1	1	1	3	neural cell adhesion molecule 2
NLGN1	22871	i	1	1	1	3	neuroligin 1
NLGN2	57555	i	1	1	1	3	neuroligin 2
NLGN3	54413	i	1	1	1	3	neuroligin 3
NLGN4X	57502	i	1	1	1	3	neuroligin 4, X-linked
NLGN4Y	22829	i	1	1	1	3	neuroligin 4, Y-linked
NRCAM	4897	i	1	1	1	3	neuronal cell adhesion molecule
NRXN1	9378	i	1	1	1	3	neurexin 1
NRXN2	9379	i	1	1	1	3	neurexin 2
NRXN3	9369	i	1	1	1	3	neurexin 3
PCDHA1	56147	i	1	1	1	3	protocadherin alpha 1
PCDHA10	56139	i	1	1	1	3	protocadherin alpha 10
PCDHA11	56138	i	1	1	1	3	protocadherin alpha 11
PCDHA2	56146	i	1	1	1	3	protocadherin alpha 2
PCDHA3	56145	i	1	1	1	3	protocadherin alpha 3
PCDHA4	56144	i	1	1	1	3	protocadherin alpha 4
PCDHA5	56143	i	1	1	1	3	protocadherin alpha 5
PCDHA6	56142	i	1	1	1	3	protocadherin alpha 6
PCDHA7	56141	i	1	1	1	3	protocadherin alpha 7
PCDHA8	56140	i	1	1	1	3	protocadherin alpha 8
PCDHAC1	56135	i	1	1	1	3	protocadherin alpha subfamily C1
PCDHAC2	56134	i	1	1	1	3	protocadherin alpha subfamily C2
PCDHB10	56126	i	1	1	1	3	protocadherin beta 10
PCDHB11	56125	i	1	1	1	3	protocadherin beta 11

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
PCDHB12	56124	i	1	1	1	3	protocadherin beta 12
PCDHB13	56123	i	1	1	1	3	protocadherin beta 13
PCDHB14	56122	i	1	1	1	3	protocadherin beta 14
PCDHB15	56121	i	1	1	1	3	protocadherin beta 15
PCDHB16	57717	i	1	1	1	3	protocadherin beta 16
PCDHB2	56133	i	1	1	1	3	protocadherin beta 2
PCDHB3	56132	i	1	1	1	3	protocadherin beta 3
PCDHB4	56131	i	1	1	1	3	protocadherin beta 4
PCDHB5	26167	i	1	1	1	3	protocadherin beta 5
PCDHB6	56130	i	1	1	1	3	protocadherin beta 6
PCDHB9	56127	i	1	1	1	3	protocadherin beta 9
PCDHGB4	8641	i	1	1	1	3	protocadherin gamma subfamily B 4
PCDHGC3	5098	i	1	1	1	3	protocadherin gamma subfamily C 3
PTPRC	5788	i	1	1	1	3	receptor type protein tyrosine phosphatase C
PTPRD	5789	i	1	1	1	3	receptor type protein tyrosine phosphatase D
PTPRJ	5795	i	1	1	1	3	receptor type protein tyrosine phosphatase J
PTPRK	5796	i	1	1	1	3	receptor type protein tyrosine phosphatase K
PTPRM	5797	i	1	1	1	3	receptor type protein tyrosine phosphatase M
PTPRT	11122	i	1	1	1	3	receptor type protein tyrosine phosphatase T
PTPRU	10076	i	1	1	1	3	receptor type protein tyrosine phosphatase U
FN1	2335	m	1	1	1	3	fibronectin 1
ITGA11	22801	m	1	1	1	3	integrin alpha 11
ITGA2	3673	m	1	1	1	3	integrin alpha 2
ITGA2B	3674	m	1	1	1	3	integrin alpha 2b
ITGA4	3676	m	1	1	1	3	integrin alpha 4
ITGA5	3678	m	1	1	1	3	integrin alpha 5
ITGA7	3679	m	1	1	1	3	integrin alpha 7
ITGA8	8516	m	1	1	1	3	integrin alpha 8
ITGA9	3680	m	1	1	1	3	integrin alpha 9
ITGAL	3683	m	1	1	1	3	integrin alpha L
ITGAM	3684	m	1	1	1	3	integrin alpha M

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
ITGAV	3685	m	1	1	1	3	integrin alpha V
ITGB1	3688	m	1	1	1	3	integrin beta 1
ITGB2	3689	m	1	1	1	3	integrin beta 2
ITGB3	3690	m	1	1	1	3	integrin beta 3
ITGB4	3691	m	1	1	1	3	integrin beta 4
ITGB6	3694	m	1	1	1	3	integrin beta 6
ITGB8	3696	m	1	1	1	3	integrin beta 8
TEK	7010	m	1	1	1	3	TEK tyrosine kinase, endothelial
THY1	7070	m	1	1	1	3	Thy-1 cell surface antigen
TNR	7143	m	1	1	1	3	tenascin R
TNXB	7148	m	1	1	1	3	tenascin XB
ROBO1	6091	ag	1	1	1	3	roundabout axon guidance receptor homolog 1
ROBO2	6092	ag	1	1	1	3	roundabout axon guidance receptor homolog 2
SEMA4D	10507	ag	1	1	1	3	semaphorin 4D
PVRL2	5819	aj	1	1	1	3	poliovirus receptor-related 2
CDON	50937	c	1	1	1	3	cell adhesion associated oncogene regulated
VCAM1	7412	c	1	1	1	3	vascular cell adhesion molecule 1
DSC2	1824	fa	1	1	1	3	desmocollin 2
DSC3	1825	fa	1	1	1	3	desmocollin 3
DSG1	1828	fa	1	1	1	3	desmoglein 1
DSG2	1829	fa	1	1	1	3	desmoglein 2
AMICA1	120425	i	0	1	1	2	adhesion molecule, interacts with CXADR antigen 1
AMIGO1	57463	i	0	1	1	2	adhesion molecule with Ig-like domain 1
AMIGO2	347902	i	0	1	1	2	adhesion molecule with Ig-like domain 2
AMIGO3	386724	i	0	1	1	2	adhesion molecule with Ig-like domain 3
ASTN1	460	i	1	0	1	2	astrotactin 1
BCAM	4059	i	0	1	1	2	basal cell adhesion molecule
BOC	91653	i	1	0	1	2	BOC cell adhesion associated oncogene regulated
CADM1	23705	i	0	1	1	2	cell adhesion molecule 1
CADM3	57863	i	0	1	1	2	cell adhesion molecule 3
CD151	977	i	0	1	1	2	CD151 molecule

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
CDH12	1010	i	1	0	1	2	cadherin 12
CDH18	1016	i	1	0	1	2	cadherin 18
CDH19	28513	i	1	0	1	2	cadherin 19
CDH20	28316	i	1	0	1	2	cadherin 20
CDH22	64405	i	1	0	1	2	cadherin 22
CDH26	60437	i	1	0	1	2	cadherin 26
CDH5	1003	i	1	0	1	2	cadherin 5
CDHR1	92211	i	1	0	1	2	cadherin-related family member 1
CDHR3	222256	i	1	0	1	2	cadherin-related family member 3
CDHR4	389118	i	1	0	1	2	cadherin-related family member 4
CELSR1	9620	i	1	0	1	2	cadherin, EGF LAG seven-pass G-type receptor 1
CELSR3	1951	i	1	0	1	2	cadherin, EGF LAG seven-pass G-type receptor 3
CERCAM	51148	i	0	1	1	2	cerebral endothelial cell adhesion molecule
CHL1	10752	i	1	0	1	2	cell adhesion molecule L1-like
CLSTN2	64084	i	1	0	1	2	calsyntenin 2
CLSTN3	9746	i	1	0	1	2	calsyntenin 3
CNTN1	1272	i	1	0	1	2	contactin 1
CNTN2	6900	i	1	0	1	2	contactin 2
CNTN3	5067	i	1	0	1	2	contactin 3
CNTN5	53942	i	1	0	1	2	contactin 5
CNTN6	27255	i	1	0	1	2	contactin 6
CTNNA1	1495	i	0	1	1	2	catenin alpha 1
CTNNA2	1496	i	0	1	1	2	catenin alpha 2
CTNNA3	29119	i	0	1	1	2	catenin alpha 3
CTNNB1	1499	i	0	1	1	2	catenin beta 1
CTNND1	1500	i	0	1	1	2	catenin delta 1
CTNND2	1501	i	0	1	1	2	catenin delta 2
DCHS2	54798	i	1	0	1	2	dachsous cadherin-related 2
DDR1	780	i	0	1	1	2	discoidin domain receptor tyrosine kinase 1
EFNA1	1942	i	0	1	1	2	ephrin A1
EFNA5	1946	i	0	1	1	2	ephrin A5

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
EFNB1	1947	i	0	1	1	2	ephrin B1
EFNB2	1948	i	0	1	1	2	ephrin B2
EMR1	2015	i	0	1	1	2	egf-like module containing mucin-like hormone receptor-like 1
EMR2	30817	i	0	1	1	2	egf-like module containing mucin-like hormone receptor-like 2
EPHA1	2041	i	1	1	0	2	eph tyrosine kinase 1
EPHA4	2043	i	1	0	1	2	EPH receptor A4
EPHB1	2047	i	1	1	0	2	eph tyrosine kinase 2
FAT3	120114	i	1	0	1	2	FAT atypical cadherin 3
FEZ1	9638	i	0	1	1	2	fasciculation and elongation protein zeta 1
FLRT1	23769	i	1	0	1	2	fibronectin leucine rich transmembrane protein 1
FLRT2	23768	i	1	0	1	2	fibronectin leucine rich transmembrane protein 2
FLRT3	23767	i	1	0	1	2	fibronectin leucine rich transmembrane protein 3
GAS6	2621	i	0	1	1	2	growth arrest-specific 6
GPR56	9289	i	0	1	1	2	G protein coupled receptor 56
ICAM1	3383	i	0	1	1	2	intercellular adhesion molecule 1
IL1RAPL1	11141	i	0	1	1	2	interleukin 1 receptor accessory protein-like 1
KIRREL2	84063	i	0	1	1	2	kin of IRRE like 2
LPHN1	22859	i	0	1	1	2	latrophilin 1
LRFN3	79414	i	1	0	1	2	leucine rich repeat and fibronectin type III domain containing 3
MAEA	10296	i	0	1	1	2	macrophage erythroblast attacher
MCAM	4162	i	1	0	1	2	melanoma cell adhesion molecule
MEGF10	84466	i	0	1	1	2	multiple EGF-like-domains 10
MEGF11	84465	i	0	1	1	2	multiple EGF-like-domains 11
MIA3	375056	i	0	1	1	2	melanoma inhibitory activity family, member 3
NCAM1	4684	i	1	0	1	2	neural cell adhesion molecule 1
NEO1	4756	i	1	0	1	2	neogenin 1
NFASC	23114	i	1	0	1	2	neurofascin
NINJ2	4815	i	0	1	1	2	ninjurin 2
NPHS1	4868	i	1	0	1	2	nephrin
NPTN	27020	i	0	1	1	2	neuroplastin

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
NRG1	3084	i	0	1	1	2	neuregulin 1
NRP2	8828	i	0	1	1	2	neuropilin 2
OPCML	4978	i	1	0	1	2	opioid binding protein
PCDH1	5097	i	1	0	1	2	protocadherin 1
PCDH10	57575	i	1	0	1	2	protocadherin 10
PCDH11X	27328	i	1	0	1	2	protocadherin 11 X-linked
PCDH11Y	83259	i	1	0	1	2	protocadherin 11 Y-linked
PCDH12	51294	i	1	0	1	2	protocadherin 12
PCDH15	65217	i	1	0	1	2	protocadherin-related 15
PCDH17	27253	i	1	0	1	2	protocadherin 17
PCDH18	54510	i	1	0	1	2	protocadherin 18
PCDH19	57526	i	1	0	1	2	protocadherin 19
PCDH20	64881	i	1	0	1	2	protocadherin 20
PCDH7	5099	i	1	0	1	2	protocadherin 7
PCDH8	5100	i	1	0	1	2	protocadherin 8
PCDH9	5101	i	1	0	1	2	protocadherin 9
PCDHA12	56137	i	1	0	1	2	protocadherin alpha 12
PCDHA13	56136	i	1	0	1	2	protocadherin alpha 13
PCDHA9	9752	i	1	0	1	2	protocadherin alpha 9
PCDHB1	29930	i	1	0	1	2	protocadherin beta 1
PCDHB18	54660	i	1	0	1	2	protocadherin beta 18
PCDHB7	56129	i	1	0	1	2	protocadherin beta 7
PCDHB8	56128	i	1	0	1	2	protocadherin beta 8
PCDHGA1	56114	i	1	0	1	2	protocadherin gamma subfamily A, 1
PCDHGA10	56106	i	1	0	1	2	protocadherin gamma subfamily A, 10
PCDHGA11	56105	i	1	0	1	2	protocadherin gamma subfamily A, 11
PCDHGA12	26025	i	1	0	1	2	protocadherin gamma subfamily A, 12
PCDHGA2	56113	i	1	0	1	2	protocadherin gamma subfamily A, 2
PCDHGA3	56112	i	1	0	1	2	protocadherin gamma subfamily A, 3
PCDHGA4	56111	i	1	0	1	2	protocadherin gamma subfamily A, 4
PCDHGA5	56110	i	1	0	1	2	protocadherin gamma subfamily A, 5

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
PCDHGA6	56109	i	1	0	1	2	protocadherin gamma subfamily A, 6
PCDHGA7	56108	i	1	0	1	2	protocadherin gamma subfamily A, 7
PCDHGA8	9708	i	1	0	1	2	protocadherin gamma subfamily A, 8
PCDHGA9	56107	i	1	0	1	2	protocadherin gamma subfamily A, 9
PCDHGB1	56104	i	1	0	1	2	protocadherin gamma subfamily B, 1
PCDHGB2	56103	i	1	0	1	2	protocadherin gamma subfamily B, 2
PCDHGB3	56102	i	1	0	1	2	protocadherin gamma subfamily B, 3
PCDHGB5	56101	i	1	0	1	2	protocadherin gamma subfamily B, 5
PCDHGB6	56100	i	1	0	1	2	protocadherin gamma subfamily B, 6
PCDHGB7	56099	i	1	0	1	2	protocadherin gamma subfamily B, 7
PCDHGC4	56098	i	1	0	1	2	protocadherin gamma subfamily C, 4
PCDHGC5	56097	i	1	0	1	2	protocadherin gamma subfamily C, 5
POSTN	10631	i	0	1	1	2	periostin
PTPRF	5792	i	1	0	1	2	receptor type protein tyrosine phosphatase F
PTPRO	5800	i	1	1	0	2	receptor-type tyrosine-protein phosphatase O
PTPRS	5802	i	1	0	1	2	receptor type protein tyrosine phosphatase S
SDK1	221935	i	1	0	1	2	sidekick cell adhesion molecule 1
SDK2	54549	i	1	0	1	2	sidekick cell adhesion molecule 2
SSPN	8082	i	0	1	1	2	sarcospan
CD36	948	m	0	1	1	2	CD36 molecule
CD44	960	m	0	1	1	2	CD44 molecule
FREM2	341640	m	1	0	1	2	FRAS1 related extracellular matrix protein 2
FREM3	166752	m	0	1	1	2	FRAS1 related extracellular matrix 3
ITGA10	8515	m	1	1	0	2	integrin alpha-10
ITGA3	3675	m	1	1	0	2	integrin alpha-3
ITGA6	3655	m	1	0	1	2	integrin alpha 6
ITGAD	3681	m	1	0	1	2	integrin alpha D
ITGAE	3682	m	1	0	1	2	integrin alpha E
ITGAX	3687	m	1	0	1	2	integrin alpha X
ITGB1BP1	9270	m	0	1	1	2	integrin beta 1 binding protein 1
ITGB3BP	23421	m	0	1	1	2	integrin beta 3 binding protein

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
ITGB5	3693	m	1	0	1	2	integrin beta 5
ITGB7	3695	m	1	0	1	2	integrin beta 7
ITGBL1	9358	m	0	1	1	2	integrin beta-like 1
KAL1	3730	m	1	0	1	2	Kallmann syndrome 1 sequence
LAMA5	3911	m	0	1	1	2	laminin alpha 5
LAMB1	3912	m	0	1	1	2	laminin beta 1
LAMC1	3915	m	0	1	1	2	laminin gamma 1
NID2	22795	m	0	1	1	2	nidogen 2
OLFM4	10562	m	0	1	1	2	olfactomedin 4
PVR	5817	m	1	0	1	2	poliovirus receptor
REG3A	5068	m	0	1	1	2	regenerating islet-derived 3 alpha
TGFBI	7045	m	0	1	1	2	transforming growth factor, beta-induced, 68kDa
THBS1	7057	m	0	1	1	2	thrombospondin 1
THBS4	7060	m	0	1	1	2	thrombospondin 4
TINAG	27283	m	0	1	1	2	tubulointerstitial nephritis antigen
TNC	3371	m	1	0	1	2	tenascin C
WNT1	7471	m	0	1	1	2	wingless-type MMTV integration site family, member 1
PLXNC1	10154	ag	0	1	1	2	plexin C1
RGMB	285704	ag	0	1	1	2	repulsive guidance molecule family member b
SEMA3E	9723	ag	1	1	0	2	semaphorin 3E
SEMA5A	9037	ag	0	1	1	2	semaphorin 5A
DLG1	1739	aj	0	1	1	2	discs large homolog 1
DLG5	9231	aj	0	1	1	2	discs large homolog 5
PVRL3	25945	aj	0	1	1	2	poliovirus receptor-related 3
PVRL4	81607	aj	1	0	1	2	poliovirus receptor-related 4
BYSL	705	c	0	1	1	2	bystin-like
CD2	914	c	0	1	1	2	CD2 molecule
CD24	1E+08	c	0	1	1	2	CD24 molecule
CD33	945	c	1	0	1	2	CD33 molecule
CD40LG	959	c	0	1	1	2	CD40 ligand
CD47	961	c	0	1	1	2	CD47 molecule

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
CD58	965	c	0	1	1	2	CD58 molecule
CD72	971	c	0	1	1	2	CD72 molecule
CD84	8832	c	0	1	1	2	CD84 molecule
CD9	928	c	0	1	1	2	CD9 molecule
CD93	22918	c	0	1	1	2	CD93 molecule
CD96	10225	c	1	0	1	2	CD96 molecule
SELE	6401	c	0	1	1	2	selectin E
SELP	6403	c	0	1	1	2	selectin P
SELPLG	6404	c	0	1	1	2	selectin P ligand
TRO	7216	c	0	1	1	2	trophinin
DSC1	1823	fa	1	0	1	2	desmocollin 1
DSG3	1830	fa	1	0	1	2	desmoglein 3
DSG4	147409	fa	1	0	1	2	desmoglein 4
LPXN	9404	fa	0	1	1	2	leupaxin
PKP1	5317	fa	0	1	1	2	plakophilin 1
PKP2	5318	fa	0	1	1	2	plakophilin 2
PKP4	8502	fa	0	1	1	2	plakophilin 4
BVES	11149	tj	0	1	1	2	blood vessel epicardial substance
CLDN1	9076	tj	0	1	1	2	claudin 1
CLDN10	9071	tj	0	1	1	2	claudin 10
CLDN11	5010	tj	0	1	1	2	claudin 11
CLDN12	9069	tj	0	1	1	2	claudin 12
CLDN14	23562	tj	0	1	1	2	claudin 14
CLDN15	24146	tj	0	1	1	2	claudin 15
CLDN16	10686	tj	0	1	1	2	claudin 16
CLDN17	26285	tj	0	1	1	2	claudin 17
CLDN18	51208	tj	0	1	1	2	claudin 18
CLDN19	149461	tj	0	1	1	2	claudin 19
CLDN2	9075	tj	0	1	1	2	claudin 2
CLDN20	49861	tj	0	1	1	2	claudin 20
CLDN22	53842	tj	0	1	1	2	claudin 22

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
CLDN23	137075	tj	0	1	1	2	claudin 23
CLDN4	1364	tj	0	1	1	2	claudin 4
CLDN5	7122	tj	0	1	1	2	claudin 5
CLDN6	9074	tj	0	1	1	2	claudin 6
CLDN7	1366	tj	0	1	1	2	claudin 7
CLDN8	9073	tj	0	1	1	2	claudin 8
CLDN9	9080	tj	0	1	1	2	claudin 9
CYTH1	9267	tj	0	1	1	2	cytohesin 1
CYTH2	9266	tj	0	1	1	2	cytohesin 2
CYTH3	9265	tj	0	1	1	2	cytohesin 3
CYTH4	27128	tj	0	1	1	2	cytohesin 4
CYTIP	9595	tj	0	1	1	2	cytohesin 1 interacting protein
JAM2	58494	tj	0	1	1	2	junctional adhesion molecule 2
JUP	3728	tj	0	1	1	2	junction plakoglobin
PVRL1	5818	tj	0	1	1	2	poliovirus receptor-related 1
ALCAM	214	i	0	0	1	1	activated leukocyte cell adhesion molecule
ASTN2	23245	i	1	0	0	1	astrotactin 2
BAI1	575	i	0	0	1	1	brain-specific angiogenesis inhibitor 1
CADM2	253559	i	0	0	1	1	cell adhesion molecule 2
CADM4	199731	i	0	0	1	1	cell adhesion molecule 4
CD200	4345	i	1	0	0	1	CD200 antigen
CD48	962	i	1	0	0	1	CD48 antigen
CD8A	925	i	1	0	0	1	CD8 antigen alpha polypeptide
CNTNAP1	8506	i	0	0	1	1	contactin associated protein 1
CNTNAP2	26047	i	0	0	1	1	contactin associated protein-like 2
CNTNAP3	79937	i	0	0	1	1	contactin associated protein-like 3
CNTNAP3B	728577	i	0	0	1	1	contactin associated protein-like 3B
CNTNAP4	85445	i	0	0	1	1	contactin associated protein-like 4
CNTNAP5	129684	i	0	0	1	1	contactin associated protein-like 5
CTNNAL1	8727	i	0	0	1	1	catenin alpha-like 1
DAB1	1600	i	0	0	1	1	disabled homolog 1

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
DCC	1630	i	1	0	0	1	deleted in colorectal cancer protein
EDIL3	10085	i	0	0	1	1	EGF-like repeats and discoidin I-like domains 3
ELFN1	392617	i	1	0	0	1	extracellular leucine-rich repeat and fibronectin type III containing 1
EPCAM	4072	i	0	1	0	1	epithelial glycoprotein-2
EPHA10	284656	i	1	0	0	1	ephrin type-A receptor 10
EPHA5	2044	i	1	0	0	1	ephrin type-A receptor 5
EPHA6	285220	i	1	0	0	1	ephrin receptor EphA6
EPHB2	2048	i	1	0	0	1	ephrin type-B receptor 2
EPHB6	2051	i	1	0	0	1	ephrin type-B receptor 6
GPR116	221395	i	1	0	0	1	G-protein coupled receptor 116
HEPACAM	220296	i	0	0	1	1	hepatic and glial cell adhesion molecule
ICAM2	3384	i	0	0	1	1	intercellular adhesion molecule 2
ICAM3	3385	i	0	0	1	1	intercellular adhesion molecule 3
ICAM4	3386	i	0	0	1	1	intercellular adhesion molecule 4
ICAM5	7087	i	0	0	1	1	intercellular adhesion molecule 5
ICOS	29851	i	0	0	1	1	inducible T-cell co-stimulator
IGSF11	152404	i	0	0	1	1	immunoglobulin superfamily, member 11
IGSF5	150084	i	0	0	1	1	immunoglobulin superfamily, member 5
IGSF9	57549	i	1	0	0	1	immunoglobulin superfamily member 9A
IGSF9B	22997	i	1	0	0	1	immunoglobulin superfamily member 9B
IL1RAP	3556	i	1	0	0	1	IL-1 receptor accessory protein
KIR2DL1	3802	i	1	0	0	1	killer cell immunoglobulin-like receptor 2DL1
KIR2DL3	3804	i	1	0	0	1	killer cell immunoglobulin-like receptor 2DL3
KIR2DL4	3805	i	1	0	0	1	killer cell immunoglobulin-like receptor 2DL4
KIR2DL5A	57292	i	1	0	0	1	killer cell immunoglobulin-like receptor 2DL5A
KIR2DL5B	553128	i	1	0	0	1	killer cell immunoglobulin-like receptor 2DL5B
KIR2DS1	3806	i	1	0	0	1	killer cell immunoglobulin-like receptor 2DS1
KIR2DS2	1E+08	i	1	0	0	1	killer cell immunoglobulin-like receptor 2DS2
KIR2DS3	3808	i	1	0	0	1	killer cell immunoglobulin-like receptor 2DS3
KIR2DS4	3809	i	1	0	0	1	killer Ig receptor|killer cell immunoglobulin-like receptor 2DS4

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
KIR2DS5	3810	i	1	0	0	1	killer cell immunoglobulin-like receptor 2DS5
KIR3DL1	3811	i	1	0	0	1	killer cell immunoglobulin-like receptor 3DL1
LRFN1	57622	i	1	0	0	1	leucine-rich repeat and fibronectin type III domain-containing protein 1
LRFN2	57497	i	1	0	0	1	leucine-rich repeat and fibronectin type-III domain-containing protein 2
LRFN4	78999	i	1	0	0	1	leucine-rich repeat and fibronectin type-III domain-containing protein 4
LRFN5	145581	i	1	0	0	1	leucine-rich repeat and fibronectin type-III domain-containing protein 5
LRRN1	57633	i	1	0	0	1	leucine-rich repeat neuronal protein 1
LRRN3	54674	i	1	0	0	1	leucine-rich repeat neuronal protein 3
LRRN4	164312	i	1	0	0	1	leucine-rich repeat neuronal protein 4
LRRN4CL	221091	i	1	0	0	1	LRRN4 C-terminal-like protein
LSAMP	4045	i	0	0	1	1	limbic system-associated membrane protein
MADCAM1	8174	i	0	0	1	1	mucosal vascular addressin cell adhesion molecule 1
MLLT4	4301	i	0	0	1	1	myeloid lymphoid or mixed-lineage leukemia translocated to 4
MPL	4352	i	1	0	0	1	thrombopoietin receptor
NINJ1	4814	i	0	0	1	1	ninjurin 1
NOTCH1	4851	i	0	1	0	1	Notch homolog 1
NRP1	8829	i	0	0	1	1	neuropilin 1
NTM	50863	i	0	0	1	1	neurotrimin
PODXL2	50512	i	0	1	0	1	podocalyxin-like protein 2
PRPH2	5961	i	0	0	1	1	peripherin 2
PRTG	283659	i	1	0	0	1	protogenin homolog
PTPRB	5787	i	1	0	0	1	receptor-type tyrosine-protein phosphatase beta
PTPRG	5793	i	1	0	0	1	receptor type protein tyrosine phosphatase gamma
PTPRH	5794	i	1	0	0	1	receptor-type tyrosine-protein phosphatase H
PTPRQ	374462	i	1	0	0	1	receptor type protein-tyrosine phosphatase Q
PTPRZ1	5803	i	1	0	0	1	receptor type protein tyrosine phosphatase zeta 1
SDC1	6382	i	1	0	0	1	syndecan 1
SDC2	6383	i	1	0	0	1	syndecan 2
SDC3	9672	i	1	0	0	1	syndecan 3

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
SDC4	6385	i	1	0	0	1	syndecan 4
ABI3BP	25890	m	1	0	0	1	ABI gene family member 3-binding protein
AGER	177	m	1	0	0	1	RAGE isoform NtRAGE-delta
ANTXR1	84168	m	0	1	0	1	tumor endothelial marker 8
AXL	558	m	1	0	0	1	tyrosine-protein kinase receptor UFO
FBLN5	10516	m	0	1	0	1	fibulin 5
FBLN7	129804	m	0	0	1	1	fibulin 7
FREM1	158326	m	1	0	0	1	FRAS1-related extracellular matrix protein 1
HSPG2	3339	m	1	0	0	1	basement membrane-specific heparan sulfate proteoglycan core protein
IBSP	3381	m	0	0	1	1	integrin-binding sialoprotein
ITGA1	3672	m	1	0	0	1	integrin alpha 1
LAMA1	284217	m	0	0	1	1	laminin alpha 1
LAMA2	3908	m	0	0	1	1	laminin alpha 2
LAMA3	3909	m	0	0	1	1	laminin alpha 3
LAMA4	3910	m	0	0	1	1	laminin alpha 4
LAMB2	3913	m	0	0	1	1	laminin beta 2
LAMB3	3914	m	0	0	1	1	laminin beta 3
LAMB4	22798	m	0	0	1	1	laminin beta 4
LAMC2	3918	m	0	0	1	1	laminin gamma 2
LAMC3	10319	m	0	0	1	1	laminin gamma 3
LYVE1	10894	m	0	1	0	1	lymphatic vessel endothelial hyaluronic acid receptor 1
MFAP4	4239	m	0	0	1	1	microfibrillar-associated protein 4
NCAN	1463	m	0	0	1	1	neurocan
RELN	5649	m	0	0	1	1	reelin
RPSA	3921	m	0	0	1	1	ribosomal protein SA
SGCE	8910	m	0	1	0	1	epsilon-sarcoglycan
THBS2	7058	m	0	0	1	1	thrombospondin 2
THBS3	7059	m	0	1	0	1	thrombospondin-3
TMEM8B	51754	m	0	1	0	1	nasopharyngeal carcinoma expressed 6
TNN	63923	m	1	0	0	1	tenascin-N
UMODL1	89766	m	1	0	0	1	olfactorin

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
USH2A	7399	m	1	0	0	1	usher syndrome type-2A protein
VCAN	1462	m	0	0	1	1	versican
WNT3A	89780	m	0	0	1	1	wingless-type MMTV integration site family member 3A
KIi0319L	79932	ag	1	0	0	1	dyslexia-associated protein KIi0319-like protein
MDGA1	266727	ag	1	0	0	1	MAM domain-containing glycosylphosphatidylinositol anchor protein 1
MDGA2	161357	ag	1	0	0	1	MAM domain-containing glycosylphosphatidylinositol anchor protein 2
NTN1	9423	ag	0	0	1	1	netrin 1
PLXNB1	5364	ag	0	1	0	1	plexin B1
PLXNB3	5365	ag	0	1	0	1	plexin B3
ROBO3	64221	ag	1	0	0	1	roundabout homolog 3
ROBO4	54538	ag	1	0	0	1	roundabout homolog 4
SEMA3G	56920	ag	1	0	0	1	semaphorin 3G
SLIT2	9353	ag	0	0	1	1	slit homolog 2
VEZT	55591	aj	0	0	1	1	vezatin
CD226	10666	c	0	0	1	1	CD226 molecule
CD6	923	c	0	0	1	1	CD6 molecule
KIR3DL3	1E+08	c	1	0	0	1	killer cell immunoglobulin-like receptor three domains long cytoplasmic tail 3
KIR3DS1	3813	c	1	0	0	1	killer cell immunoglobulin-like receptor 3DS1
PECAM1	5175	c	0	0	1	1	platelet endothelial cell adhesion molecule 1
SELL	6402	c	0	0	1	1	selectin L
SPAM1	6677	c	0	0	1	1	sperm adhesion molecule 1
FERMT2	10979	fa	0	1	0	1	fermitin family homolog 2
LIMS2	55679	fa	0	0	1	1	LIM and senescent cell antigen-like domains 2
LPP	4026	fa	0	0	1	1	LIM domain containing preferred translocation partner in lipoma
NEDD9	4739	fa	0	0	1	1	neural precursor cell expressed, developmentally down-regulated 9
PEAK1	79834	fa	0	1	0	1	NKF3 kinase family member
PKP3	11187	fa	0	0	1	1	plakophilin 3
TSC1	7248	fa	0	1	0	1	tuberous sclerosis 1 protein
ZYX	7791	fa	0	0	1	1	zyxin

symbol	#entrezID	CAM type	Inter Pro	GO	NCBI	Sum	name
MAG	4099	my	0	0	1	1	myelin associated glycoprotein
OMG	4974	my	0	0	1	1	oligodendrocyte myelin glycoprotein
AJAP1	55966	tj	0	0	1	1	adherens junctions associated protein 1
AJUBA	84962	tj	0	0	1	1	ajuba LIM protein
CLDN3	1365	tj	0	0	1	1	claudin 3
DSP	1832	tj	0	0	1	1	desmoplakin
DST	667	tj	0	0	1	1	dystonin
ESAM	90952	tj	0	0	1	1	endothelial cell adhesion molecule
F11R	50848	tj	0	0	1	1	F11 receptor
JAM3	83700	tj	0	0	1	1	junctional adhesion molecule 3

Columns: Gene symbol, ENTREZ gene ID number, annotated CAM type (see text), sources that identify this CAM and sum of the number of sources, and gene name. Genes are sorted by number of sources, CAM type, then alphabetically by gene. CAM types: i: information predominant CAM, m: primarily involved in interactions with cell matrix, ag: primary roles in axonal guidance, aj: primary role in adherens junctions, c: primary roles in cell/cell interactions, principally in immune system, fa: primary involvement in focal adhesions, tj: primary involvement in tight junctions. Please note likely involvement of many of the products of these CAM genes in multiple functions (*esp* cadherins).

Ann N Y Acad Sci. Author manuscript; available in PMC 2016 September 01.

Table II

Likely CAMs for which single nucleotide polymorphisms are likely to provide partial or complete human “knockout” (79).

					frequency		
gene	**chr:bp**	**Ref**	**Alt**	**Af**	**EuAm**	**As**	**Eu**
CDH19	18:64235709	C	A	0	0.28	0	0
CDHR2	5:175998270	C	T	0	0	0.2	0
CEACAM1	19:43031354	G	A	0	0.28	0	0
CLDN20	6:155597004	G	A	0	0	0.2	0.8
CNTN6	3:1415319	G	T	0	0	0	0.1
DSG4	18:28993525	C	T	0.4	0	0	0
	18:28979436	T	G	0	0	0.2	0
	18:28993216	C	T	0	0	0	0.1
DSP	6:7584223	T	A	0	0.28	0	0
	6:7583371	G	A	0	0	0.2	0
EPHB1	3:134885826	C	T	0	0.28	0	0
FAT1	4:187518227	C	A	0.2	0	0	0
	4:187630007	C	T	0.2	0	0	0
FAT2	5:150923303	G	A	0	0.28	0	0
ICAM3	19:10446525	C	T	0	0.28	0	0
	19:10445374	G	A	0	0	0.2	0
ITGA10	1:145542272	C	T	0	0.28	0	0.1
	1:145539790	C	T	0	0	0.2	0
	1:145533485	C	T	0	0	0	0.1
ITGA11	15:68641223	C	A	0	0	0	0.1
ITGA7	12:56092316	G	A	0	0	0.2	0
ITGAD	16:31425869	C	T	0	0	0	0.1
ITGB6	2:160994697	C	A	0	0.28	0	0
	2:160993947	C	A	0.2	0	0	0
KIR2DL1	19:55286866	C	G	0	0	0	0.5
	19:55331238	C	T	0	0	0.2	0
LAMA2	6:129475705	A	T	0	0	0.5	0
	6:129636987	C	T	0	0	0	0.1

					frequency		
gene	**chr:bp**	**Ref**	**Alt**	**Af**	**EuAm**	**As**	**Eu**
LAMC3	9:133954628	C	T	0.2	0	0	0
LRRN3	7:110764535	C	T	0	0	0.2	0
OLFM4	13:53617308	C	T	0	0	0	2.6
	13:53624743	T	A	0	0	0.2	0
PCDH18	4:138449706	G	A	0	0.28	0	0
PCDH7	4:30725762	A	T	0	0	0	0.1
PCDHB1	5:140432438	C	T	0	0	0.2	0
PCDHB4	5:140503823	C	G	0	0	0	0.3
PCDHB7	5:140554309	G	T	0	1.66	0.2	0.3
	5:140553499	G	T	0.4	0	0	0
PCDHB8	5:140558148	T	G	0.2	0	0	0
	5:140558623	C	T	0	0	0.2	0
	5:140559627	C	A	0	0	0.2	0
PCDHGA10	5:140794929	C	T	0.2	1.1	0	0.4
PCDHGA8	5:140773284	C	G	0	0	0	0.1
PCDHGA9	5:140783659	C	T	0	0	0.5	0
PCDHGC5	5:140869257	C	T	0.2	0	0	0
	5:140870361	C	T	0	0	0	0.1
	5:140870415	C	T	0	0	0	0.1
PTPRH	19:55697711	G	A	0	0.28	0	0.5
PTPRU	1:29650188	G	A	0.2	0	0	0
ROBO2	3:77611841	C	A	0.2	0	0	0
SDK1	7:4213906	C	A	0	0	0.2	0
SELL	1:169676573	C	T	0.2	0	0	0
SEMA5A	5:9197325	G	T	0	0	0.2	0
SPAM1	7:123599667	G	A	0	0	0	0.1
THBS4	5:79372855	C	T	0.2	0	0	0
	5:79366570	C	T	0	0	0	0.1
TNN	1:175052986	C	T	0	0	0.2	0

Columns list: gene symbol, chromosome and basepair of variant, major/reference allele, minor/mutant allele, and minor allele frequencies in 1000 genomes data from African, US individuals of European ancestry, Asian and European samples. Note that many CAM genes are also sites for copy number and other variation that can also provide knockouts.

Table III

Candidate addiction-related CAMs. Genes listed both a) encode "likely" *bona fide* human cell adhesion molecules and b) are identified by at least 3 independent case *vs* control GWAS studies of addiction-related phenotypes by clusters of SNPs with $10^{-2} > p > 10^{-8}$. Genes are arranged by the number of addiction-related case-control sample pairs in which they are identified, then alphabetical order.

gene	type	#samp	structure	ec motifs
CDH13	iCAM	13	GPI	Cdh
CSMD1	iCAM	12	1TM	Cub/sushi
PTPRD	iCAM	12	1TM	Ig/Fn
CLSTN2	iCAM	10	1TM	Cdh/laminin
DAB1	iCAM	10	cyt/sec	pTyr-bind dom
CNTNAP2	iCAM	8	1TM	EGF/laminin
CTNNA2	iCAM	8	cyt	Vinculin
PTPRM	iCAM	8	1TM	Ig/Fn/MAM
ASTN1	iCAM	7	1TM	Mem-attack comp/Fn/EGF
CNTNAP5	iCAM	7	1TM	EGF/laminin
CTNNA3	aj	7	cyt	Vinculin
DSCAM	iCAM	7	1TM	Ig/Fn
NRG1	iCAM	7	1TM	Ig/EGF
OPCML	iCAM	7	GPI	Ig
CHL1	iCAM	6	1TM	Ig/Fn
CNTN4	iCAM	6	GPI	Ig/Fn
CNTN5	iCAM	6	GPI	Ig/Fn
CTNND2	iCAM	6	cyt	Arm/β-catenin-l
EPHB1	iCAM	6	1TM	SAM/Fn/TNFR
LAMA1	m	6	EC	Lam
NLGN1	iCAM	6	1TM	Esterase
NRXN3	iCAM	6	1TM	EGF/laminin
PTPRT	iCAM	6	1TM	Ig/Fn/MAM
USH2A	m	6	EC	EGF/lam/Fn
CDH11	iCAM	5	1TM	Cdh
CSMD2	iCAM	5	1TM	Cub/sushi

gene	type	#samp	structure	ec motifs
ITGA9	m	5	TM-assoc	Int
ITGB8	m	5	1TM	Int
LAMA2	m	5	EC	EGF/lam
MDGA2	ag	5	GPI	Ig/MAM
PLXNC1	ag	5	1TM	Sema/plexin
TEK	m	5	1TM	Ig/Fn/EGF
ASTN2	iCAM	4	1TM	Mem-attack comp/Fn/EGF
CDH2	iCAM	4	1TM	Cdh
CDH4	iCAM	4	1TM	Cdh
CDH6	iCAM	4	1TM	Cdh
CDH7	iCAM	4	1TM	Cdh
DSCAML1	iCAM	4	1TM	Ig/Fn
FAT3	iCAM	4	1TM	Cdh/EGF/lam
FLRT2	iCAM	4	1TM	LRR/Fn
FREM1	m	4	EC	C lectin/Calxbeta/CSPG
FREM2	m	4	1TM	C lectin/Calxbeta/CSPG
ITGA1	m	4	TM-assoc	Int
LPP	fa	4	cyt/nuc	LIM/GKAP
LRRN2	iCAM	4	1TM	LRR/Ig
NCAM1	iCAM	4	1TM	Ig/Fn
PCDH15	iCAM	4	1TM	Cdh
PCDH9	iCAM	4	1TM	Cdh
PTPRB	iCAM	4	1TM	Fn
PTPRG	iCAM	4	1TM	Fn
PTPRK	iCAM	4	1TM	Ig/Fn/MAM
RELN	m	4	EC	Rln/EGF
ROBO2	ag	4	1TM	Ig/Fn
SELE	c	4	1TM	CCP/Clect/EGF
SELL	c	4	1TM	CCP/Clect/EGF
SEMA5A	ag	4	1TM	Thrombosp/Plxn/Sema
TNR	m	4	EC	Fn/FReD/EGF

gene	type	#samp	structure	ec motifs
AJAP1	tj	3	1TM	AJAP1/PANP C-term
ANTXR1	m	3	1TM	Anth_Ig/vWA_ATR
CDH12	iCAM	3	1TM	Cdh
CDH18	iCAM	3	1TM	Cdh
CDHR3	iCAM	3	1TM	Cdh
CELSR1	iCAM	3	7TM	Cdh/EGF/lam
DCC	iCAM	3	1TM	Ig/Fn
DSC3	fa	3	1TM	Cdh
DSG4	fa	3	1TM	Cdh
DSP	tj	3	cyt	PLEC/SPEC/ApoLp-III_like
EFNA5	iCAM	3	GPI	Ephrin-A
FREM3	m	3	EC	Calxbeta/CSPG
GPR116	iCAM	3	7TM	Ig/SEA/lathrophilin
ITGA6	m	3	TM-assoc	Int
ITGBL1	m	3	1TM	Int
LAMC2	m	3	EC	Lam/EGF
LAMC3	m	3	EC	Lam/EGF
LRFN5	iCAM	3	1TM	LRR/Fn/Ig
LSAMP	iCAM	3	GPI	Ig
MEGF11	iCAM	3	1TM	EGF
NFASC	iCAM	3	1TM	Ig/Fn
NINJ2	iCAM	3	1TM	Ninj
NRP1	iCAM	3	1TM	MAM/Cub/FA58C
SDK1	iCAM	3	1TM	Ig/Fn
SELP	c	3	1TM	CCP/Clect/EGF

Columns: Gene symbol, CAM type, number of samples out of 15 total that identify this gene, structural elements of protein, protein motifs. CAM types: iCAM: information predominant CAM, m: primarily involved in interactions with cell matrix, ag: primary roles in axonal guidance, aj: primary role in adherens junctions, c: primary roles in cell/cell interactions, principally in immune system, fa: primary involvement in focal adhesions, tj: primary involvement in tight junctions. Structural elements: GPI: glycosylphosphatidyl inositol modified, cyt: cytoplasmic; sec: secreted; EC: extracellular. Motifs: Cdh: cadherin; Ig: immunoglobulin; Fn: fibronectin; lam: laminin; EGF epidermal growth factor; pTyr bind dom: phosphor tyrosine binding domain; MAM: ; Mem-attack comp: membrane attack complex; Armad b-catenin-l: Armadillo and beta catenin like; SAM: sterile alpha motif; TNFR: tumor necrosis factor receptor; Int: integrin; sema: semaphorin domain; LRR: leucine-rich repeat; C lectin: C-type lectin; Calxbeta: Calx beta domain; CSPG: chondroitin sulfate proteoglycan; LIM: LIM interacting protein; GKAP: guanylate-kinase-associated protein;CCP: complement control protein; Thrombosp: thrombospondin; FReD: Fibrinogen-related domains. Samples for addiction- or quit success-related GWAS are described in (21,33,35–39,41,42,44,80,81); more details of analyses are available on request.