

Databases and ontologies

# Mapping genes for calcium signaling and their associated human genetic disorders

Matthias Hörtenhuber<sup>1</sup>, Enrique M. Toledo<sup>1</sup>, Erik Smedler<sup>1</sup>, Ernest Arenas<sup>1</sup>, Seth Malmersjö<sup>2</sup>, Lauri Louhivuori<sup>1</sup> and Per Uhlén<sup>1,\*</sup>

<sup>1</sup>Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-171 77 Stockholm, Sweden and

<sup>2</sup>Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA 94305, USA

\*To whom correspondence should be addressed.

Associate Editor: John Hancock

Received on May 2, 2016; revised on April 7, 2017; editorial decision on April 12, 2017; accepted on April 18, 2017

## Abstract

**Motivation:** Signal transduction via calcium ions ( $\text{Ca}^{2+}$ ) represents a fundamental signaling pathway in all eukaryotic cells. A large portion of the human genome encodes proteins used to assemble signaling systems that can transduce signals with diverse spatial and temporal dynamics.

**Results:** Here, we provide a map of all of the genes involved in  $\text{Ca}^{2+}$  signaling and link these genes to human genetic disorders. Using Gene Ontology terms and genome databases, 1805 genes were identified as regulators or targets of intracellular  $\text{Ca}^{2+}$  signals. Associating these 1805 genes with human genetic disorders uncovered 1470 diseases with mutated 'Ca<sup>2+</sup> genes'. A network with scale-free properties appeared when the  $\text{Ca}^{2+}$  genes were mapped to their associated genetic disorders.

**Availability and Implementation:** The  $\text{Ca}^{2+}$  genome database is freely available at <http://cagedb.uhlenlab.org> and will foster studies of gene functions and genetic disorders associated with  $\text{Ca}^{2+}$  signaling.

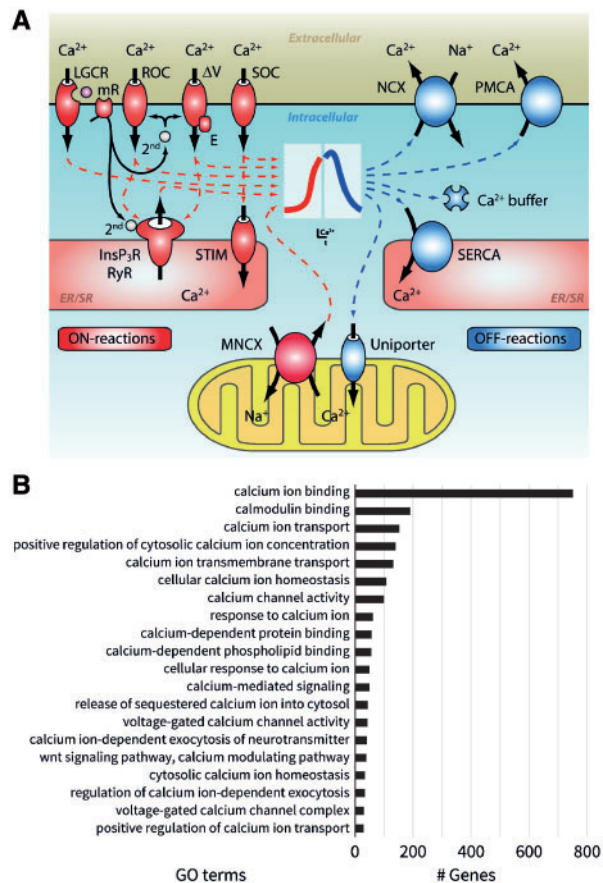
**Contact:** per.uhlen@ki.se

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

## 1 Introduction

The calcium ion ( $\text{Ca}^{2+}$ ) is a universal cell-signaling messenger with diverse roles in a wide range of biological processes and human diseases (Berridge, 2012; Clapham, 2007; Uhlén and Fritz, 2010). A large number of  $\text{Ca}^{2+}$ -channels,  $\text{Ca}^{2+}$ -pumps and organelles are responsible for the strict regulation of intracellular  $\text{Ca}^{2+}$  concentrations (Fig. 1A) (Berridge *et al.*, 2003; Bootman *et al.*, 2009; Uhlén and Fritz, 2010) because high  $\text{Ca}^{2+}$  levels are toxic to the cell (Zhivotovsky and Orrenius, 2011). To date, more than 170 000 publications in NCBI's PubMed include  $\text{Ca}^{2+}$  in the title, thereby demonstrating the diverse role of  $\text{Ca}^{2+}$  in biological processes and the immense interest among scientists to understand the physiological and pathological processes regulated by  $\text{Ca}^{2+}$ . However, this large quantity of available literature also makes it difficult to establish connections between disparate discoveries made over previous decades using diverse model systems.

Dysfunctional  $\text{Ca}^{2+}$  signaling has been reported in a large variety of human genetic disorders (Bidaud *et al.*, 2006; Chakraborti *et al.*, 2007; Delmas *et al.*, 2004; Giacomello *et al.*, 2013; Nedergaard *et al.*, 2010; Rajakulendran *et al.*, 2012; Ramasamy, 2008; Uhlén *et al.*, 2006). Given that  $\text{Ca}^{2+}$  regulates a myriad of physiological systems at every level, including the modulation of the cell's membrane potential and underlying channels and ion transporters, kinases and transcription factors, signaling cascades that regulate gene expression, as well as tissue network activity via gap junctions and peptide/transmitter release to name a few, it should not be surprising that disruptions of  $\text{Ca}^{2+}$  homeostasis both extra- and intracellularly, underlie a host of diseases in virtually every organ system in the human body. For example,  $\text{Ca}^{2+}$  signaling plays a decisive role in the blood coagulation pathway (Triplett, 2000), a critical role in muscle contraction (smooth, striated and cardiac) with many skeletal myopathies resulting from intracellular  $\text{Ca}^{2+}$  dysregulation



**Fig. 1.** Regulating the cytosolic  $\text{Ca}^{2+}$  level. **(A)** Schematic diagram illustrating the main regulators of the cytosolic  $\text{Ca}^{2+}$  level. Cytosolic  $\text{Ca}^{2+}$  signals are generated through the concerted action of cellular mechanisms that increase (ON-reactions, red) and decrease (OFF-reactions, blue) the concentration of  $\text{Ca}^{2+}$  in the cytoplasm. Typical  $\text{Ca}^{2+}$  signals are initiated by stimuli that trigger the entry of external  $\text{Ca}^{2+}$  through ligand-gated channel receptors (LGCR), receptor operated channels (ROC) or voltage ( $\Delta V$ ) gated  $\text{Ca}^{2+}$  channels in the plasma membrane or by the activation of metabotropic receptors (mR) that stimulate PLC- and  $\text{InsP}_3$ -mediated  $\text{Ca}^{2+}$ -release from the ER/SR and the accompanying refilling of  $\text{Ca}^{2+}$  stores via store operated channels (SOC). When the cytosolic level of  $\text{Ca}^{2+}$  increases,  $\text{Ca}^{2+}$  itself stimulates  $\text{InsP}_3\text{R}$ s and/or  $\text{RyR}$ s to release further  $\text{Ca}^{2+}$  into the cytoplasm. During this phase,  $\text{Ca}^{2+}$  buffers bind  $\text{Ca}^{2+}$ , which contributes to the decrease in the cytosolic concentration of free  $\text{Ca}^{2+}$ . When the  $\text{Ca}^{2+}$  concentration reaches high levels, the plasma membrane  $\text{Ca}^{2+}$ -ATPase (PMCA) and  $\text{Na}^+/\text{Ca}^{2+}$ -exchanger (NCX) extrude  $\text{Ca}^{2+}$  to the outside, whereas the ER/SR  $\text{Ca}^{2+}$ -ATPase (SERCA) pumps  $\text{Ca}^{2+}$  back into the ER/SR. High levels of  $\text{Ca}^{2+}$  is also regulated by the mitochondria via the mitochondrial NCX (MNCX) and the  $\text{Ca}^{2+}$  uniporter. Additionally second messengers (2nd) and regulatory proteins (E) are able to modulate intracellular  $\text{Ca}^{2+}$ . **(B)** Bar chart of 20 annotated GO terms associated with the most  $\text{Ca}^{2+}$  genes

(Hernandez-Ochoa *et al.*, 2015), and clear links in  $\text{Ca}^{2+}$ -handling gene mutations to inherited arrhythmia syndromes (Burashnikov *et al.*, 2010; Priori *et al.*, 2001). Moreover, autosomal dominant polycystic kidney disease, the most common inherited pathology of the kidneys, is caused by the mutation of two genes (PKD1 and PKD2) that encode integral membrane proteins that can transport  $\text{Ca}^{2+}$  (Kuo and Ehrlich, 2012). In neurodegenerative diseases,  $\text{Ca}^{2+}$  signaling has been implicated in conditions as diverse as amyotrophic lateral sclerosis (Rowland and Schneider, 2001), Parkinson's disease (Thomas and Beal, 2007), skeletal muscle atrophy (Jackman and Kandarian, 2004), familial frontotemporal dementia (Furukawa

*et al.*, 2003), Alzheimer's disease (Bezprozvanny and Mattson, 2008), scrapie (Kristensson *et al.*, 1993), progressive supranuclear palsy (Zemaitaitis *et al.*, 2000), Tourette syndrome (Martino *et al.*, 2009), Huntington's disease (Bezprozvanny and Hayden, 2004) and Wolfram syndrome (Osman *et al.*, 2003). Further understanding of how these and other diseases are interlinked through  $\text{Ca}^{2+}$  signaling is important for understanding the evolution of genetic disorders and developing more efficient treatment strategies in the future.

Here, we present a map of genes involved in  $\text{Ca}^{2+}$  signaling and the human genetic disorders associated with these genes. We provide an online database (<http://cagedb.uhlenlab.org>) that is automatically updated to incorporate novel discoveries related to gene functions and genetic diseases. Clustering genetic disorders with regard to common  $\text{Ca}^{2+}$  genes reveals interesting and unforeseen connections between human diseases. The  $\text{Ca}^{2+}$  genome and its associated online database will be valuable resources that should help facilitate a better understanding of the connection between  $\text{Ca}^{2+}$  signaling genes and associated diseases.

## 2 Materials and methods

### 2.1 Software

The main script was written in Ruby version 2.2.2 (<https://www.ruby-lang.org>). Data files from the queries were saved in SQLite version 3.7.14.1. The CaGeDB software tool can be downloaded from <http://cagedb.uhlenlab.org>.

### 2.2 Databases

Six different databases (Gene Ontology, Ensembl, NCBI Gene, OMIM, Comparative Toxicogenomic Database and Medical Subject Headings Database (MeSH)) were used in this study. A Ruby script automatically executed each query, and the results were saved in an SQLite database. A description of the databases used in each step is provided below. The data presented in this work were updated on February 19, 2017.

### 2.3 Genes associated with $\text{Ca}^{2+}$

Gene ontology (GO, <http://geneontology.org> (Ashburner *et al.*, 2000)) terms associated with  $\text{Ca}^{2+}$  signaling were determined by searching the GO names for text-strings related to  $\text{Ca}^{2+}$ . The resulting list of GO terms was manually filtered to ensure true  $\text{Ca}^{2+}$ -associations. The evidence codes associated with each GO annotation were not used in this study.

Two gene databases, Ensembl (<http://www.ensembl.org>) and NCBI Gene (<http://www.ncbi.nlm.nih.gov/gene>), were queried to search for genes classified with a  $\text{Ca}^{2+}$  signaling-related GO term from our list. We applied both Ensembl version 87 and the human genome assembly GRCh38.p7. The GO term-based queries were merged first according to their HGNC (HUGO (Human Genome Organization) Gene Nomenclature Committee (HGNC, <http://www.genenames.org>)) gene symbol and then according to their NCBI ID using mapping between the two database IDs provided by Ensembl. All of the query data files were saved in an SQLite database.

### 2.4 Genetic disorders related to $\text{Ca}^{2+}$

Human diseases associated with the  $\text{Ca}^{2+}$  genome were determined by querying the Comparative Toxicogenomic Database (CTD, <http://ctdbase.org> (Davis *et al.*, 2015), from the Department of Bioinformatics at the MDI Biological Laboratory, ME, U.S. and the Department of Biological Sciences at the North Carolina State

University, NC, U.S.), which uses the online database Mendelian Inheritance in Man (OMIM, <http://omim.org>, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, MD, U.S.). Gene IDs used to query diseases were unique genes hits associated with the curated calcium GO-term ontology list and therefore the results of the disease query are independent on how many GO annotations a gene ID contained. The terms of the genetic disorders obtained from the OMIM database via the CTD were thereafter categorized using the Medical Subject Headings (MeSH, <http://www.nlm.nih.gov/mesh/>) database (Davis *et al.*, 2012).

The Ca<sup>2+</sup>-related genes and their associated diseases were represented as an adjacency matrix, and the distances between diseases were calculated as Pearson correlations. The numbers of clusters were determinate empirically with the ConsensusCluster Plus R package (Wilkerson and Hayes, 2010). Cluster stability was calculated by multiscale bootstrapping resampling (Suzuki and Shimodaira, 2006). A term frequency analysis per cluster was performed after the removal of numbers, punctuation and the words disease(s), disorder(s), with, and, syndrome, type and cell. The heat map representation shows term frequency by cluster and filtered by a variance of 90%. The analysis was performed using R software.

## 2.5 Graphical presentations and statistical analyses

The sunburst diagrams, force-directed graphs and network plots were generated using Gephi 0.8.2. The regression analyses and correlation coefficients were calculated using MATLAB.

## 3 Results

### 3.1 Identification of Ca<sup>2+</sup> genes

To map genes related to Ca<sup>2+</sup> signaling, we took advantage of Gene Ontology (GO) terms, established by the Gene Ontology Consortium (Gene Ontology, 2008). The GO project has developed three structured ontologies that describe gene products in a species-independent manner in terms of their associated biological processes, cellular components and molecular functions. We generated a list of GO terms by searching the entire database for strings matching entries related to Ca<sup>2+</sup> signaling. The list was then manually checked to filter out entries exclusively involving Ca<sup>2+</sup> signaling, which resulted in a total of 241 GO terms (Supplementary Table S1). Among these GO terms, 161 belonged to biological processes, 65 belonged to molecular functions and 15 belonged to cellular components. The large bias toward biological processes reflects the diverse role of Ca<sup>2+</sup> signaling in regulating a wide variety of cellular processes (Berridge *et al.*, 2000).

All of the GO term annotations have a specific evidence code that reflects the type of work or analysis described in the cited reference that associates the gene to the GO term. Approximately three quarters (72%) of the identified Ca<sup>2+</sup> signaling genes were annotated automatically without human curation, i.e. evidence code 'Inferred from Electronic Annotation (IEA)'. This is a relatively low proportion because 98% of all GO annotations are based on evidence code IEA (du Plessis *et al.*, 2011). The other evidence codes are more equally distributed among annotations from experimental data and non-experimental data, and exact definitions of all of the evidence codes can be found elsewhere (<http://geneontology.org/page/guide-go-evidence-codes>).

We identified all of the human genes associated with the list of 241 Ca<sup>2+</sup> signaling GO terms by querying two different gene databases, Ensembl and NCBI Gene. The Ensembl database contained

1739 genes that matched our list of GO terms, whereas the NCBI Gene database contained 1575 genes. Merging the two queries together resulted in a list of 1805 unique genes involved in Ca<sup>2+</sup> signaling. Among these genes, 1668 had well-defined gene symbols. The GO terms associated with the highest number of Ca<sup>2+</sup> signaling genes are shown in Figure 1B, and the largest term 'calcium ion binding' contained 751 genes. Altogether, the top 50 GO terms accounted for 4242 genes, thereby indicating that one gene can be linked to several GO terms.

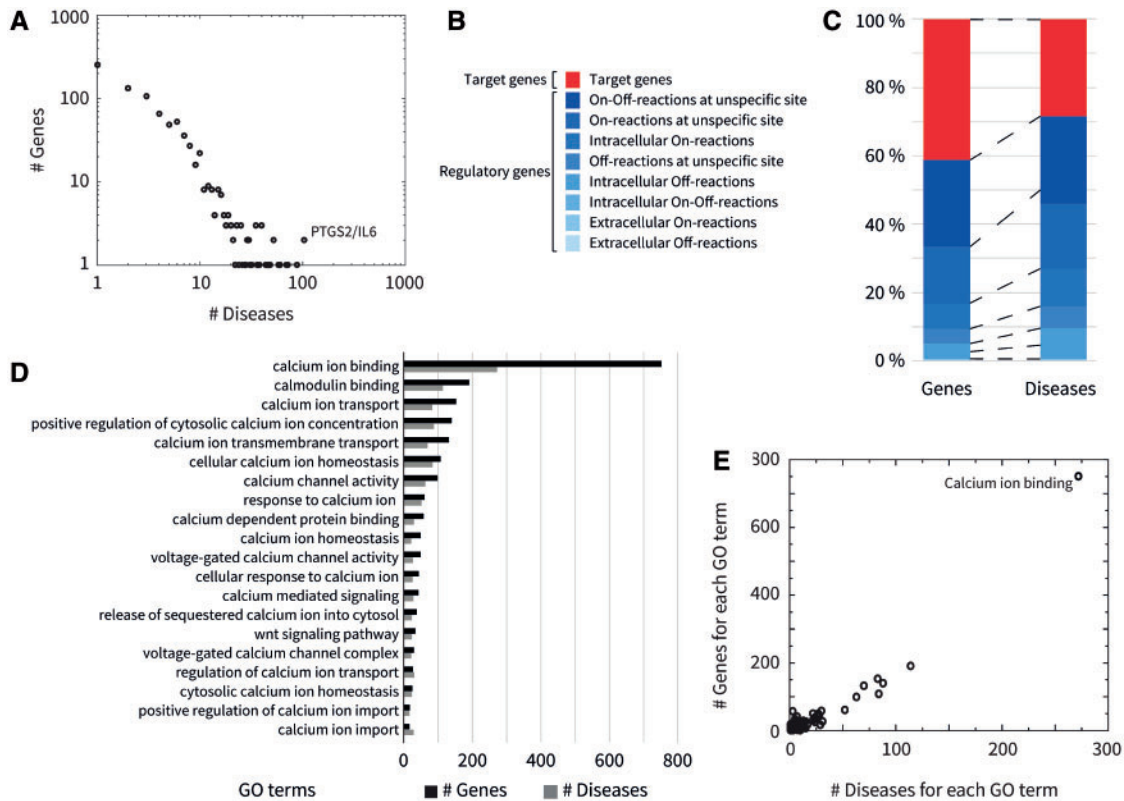
### 3.2 Mapping Ca<sup>2+</sup> genes to human genetic disorders

Next, we sought to map human genetic disorders associated with the Ca<sup>2+</sup> genome. For this purpose, we used the open Online Mendelian Inheritance in Man database (OMIM, <http://omim.org>) via the Comparative Toxicogenomic Database (CTD, <http://ctdbase.org>). Among the 1805 mapped Ca<sup>2+</sup> genes, 912 genes were associated with one or several genetic disorders. Searching the OMIM database resulted in a total number of 5205 disease-gene associations, with 1470 different genetic disorders as a result of a direct mutation in a Ca<sup>2+</sup> gene, demonstrating that one specific gene can be linked to multiple diseases (or subtypes of a given disease).

The number of diseases associated with Ca<sup>2+</sup> genes followed a power distribution (Fig. 2A). Datasets that have a large variance, which is the case for power distributions, are typical for scale-free networks (Malmersjo *et al.*, 2013; Smedler *et al.*, 2014). For example, PTGS2 (prostaglandin-endoperoxide synthase 2, also known as COX2) and IL6 (interleukin 6) were connected to approximately 100 diseases each (Fig. 2A), whereas CACNA1B (calcium channel, voltage-dependent, N-type, alpha 1B subunit) and TPT1 (tumor protein, translationally controlled 1) were only connected to one disease. Interestingly, the Ca<sup>2+</sup> genome contained 21 genes that accounted for one quarter of all the gene-disease associations.

To determine the specific role of individual genes involved in Ca<sup>2+</sup> signaling, we divided the GO terms into two major classes: 'Target genes' and 'Regulatory genes' (Fig. 2B). 'Target genes' typically respond to changes in the intracellular Ca<sup>2+</sup> concentration, such as kinases and transcription factors. 'Regulatory genes' are defined as genes that regulate the intracellular Ca<sup>2+</sup> concentration, such as Ca<sup>2+</sup> channels, Ca<sup>2+</sup> pumps and Ca<sup>2+</sup> binding proteins (including Ca<sup>2+</sup> buffer proteins). 'Regulatory genes' were further classified with regards to their regulatory function, where increases in free cytosolic Ca<sup>2+</sup> are termed 'ON-reactions' and decreases are termed 'OFF-reactions' as described by Michael Berridge (Berridge *et al.*, 2000). We also classified the GO terms belonging to 'Regulatory genes' according to their cellular location (Fig. 2B). The schematic diagram in Figure 1A shows the typical proteins encoded by 'Regulatory genes' that are responsible for both 'ON-reactions' and 'OFF-reactions'.

Interestingly, when the Ca<sup>2+</sup>-related diseases were divided into our two main categories, we found that 'Regulatory genes' accounted for more than 70% of the total number of diseases (Fig. 2C), which demonstrates that perturbations in the Ca<sup>2+</sup> signaling machinery can generate a large number of genetic disorders. Furthermore, one quarter of all of the gene-disease associations involved genes related to increases in cytosolic Ca<sup>2+</sup> concentrations. The 20 GO-terms with the largest number of diseases are listed in Figure 2D, with 'calcium ion binding' number one. When the entire list of GO terms was plotted using the number of genes and the number of diseases as axes, the terms appeared correlated (Fig. 2E). Performing a linear regression analysis showed a strong correlation ( $R = 0.95707$ ) among the GO terms.



**Fig. 2.** GO term categories of annotated Ca<sup>2+</sup> genes and their diseases. **(A)** Log-log plot of the distribution of diseases to genes associated with Ca<sup>2+</sup>. **(B)** Categorizing Ca<sup>2+</sup>-related GO terms into Target genes and Regulatory genes. **(C)** Percentage of Ca<sup>2+</sup>-related GO term categories with respect to genes (left) and genetic disorders (right). **(D)** Bar chart of the 20 annotated GO terms associated with the most Ca<sup>2+</sup> genes plus genetic disorders. **(E)** Ca<sup>2+</sup>-related GO terms plotted as the number of genes versus the number of diseases

### 3.3 Analyzing diseases linked to Ca<sup>2+</sup> signaling

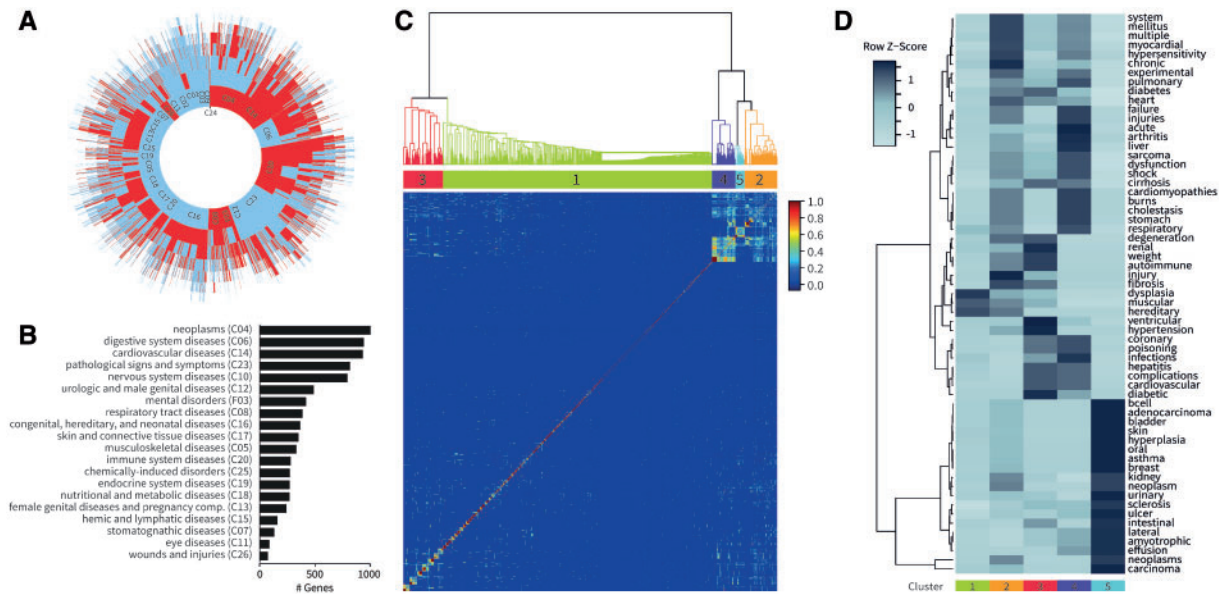
We then matched the genetic disorders queried from the OMIM database to the Medical Subject Headings (MeSH), which is the National Library of Medicine's controlled vocabulary thesaurus. MeSH descriptors are organized in 16 categories, with Diseases comprising one category. Each category is further divided into sub-categories, in which descriptors are arrayed hierarchically from most general to most specific for up to 13 hierarchical levels. For example the subcategory 'Pathological Conditions, Signs, and Symptoms' has a scope 'Abnormal anatomical or physiological conditions and objective or subjective manifestations of disease, not classified as disease or syndrome' with up to 793 associated descriptors, i.e. symptoms associated with edema, heart murmurs, renal colic, fatigue, etc. (For a full description of MeSH vocabulary go to the MeSH website: <https://www.nlm.nih.gov/mesh>). The structure of the branching hierarchy frequently represents a compromise among the views and needs of particular disciplines and users, in the absence of any single universally accepted arrangement. As such, each MeSH descriptor appears in at least one place, and may appear in as many additional places as may be appropriate (The complexity of the branching is apparent in Figure 3A and Supplementary Fig. S1). Out of the entire list of 11 721 MeSH terms describing human diseases, 13% (1470 terms) had at least one reported association with a gene related to Ca<sup>2+</sup> signaling (Fig. 3A). The category of 'neoplasms (C04)' i.e. abnormal growth in a body, such as tumors, was the most common main category, and 975 gene connections were observed between the category and its sub-categories (Fig. 3B). Remarkably, Ca<sup>2+</sup> associated diseases encompass at least one sub-

category in every branch of the MeSH tree hierarchy, except for the parent category 'Disorders caused by external forces rather than by physiologic dysfunction or by pathogens (C21)'. The sub-category with the most associated genes was 'prostatic neoplasms' (prostate cancer), which presented 74 gene connections.

When the relationships between GO terms and gene diseases were plotted with a force-directed graph (Supplementary Fig. S1), properties of scale-free network became apparent. Several large nodes originating from GO terms (pink dots) were observed with more than 500 genes (yellow dots) for neoplasms (MeSH ID C04), cardiovascular diseases (MeSH ID C14), digestive system diseases (MeSH ID C06), nervous system diseases (MeSH ID C10) and pathological conditions, signs and symptoms (MeSH ID C23). Similarly, the number of diseases (red dots in Supplementary Fig. S1) connected to each gene varied greatly.

The Comparative Toxicogenomic Database disease vocabulary (<http://ctdbase.org>) uses modified descriptors from MeSH combined with genetic disorders from OMIM to map OMIM diseases within the hierarchical MeSH disease vocabulary, thus expanding disease representation. When organizing the top five Ca<sup>2+</sup> genes based on the number of associated diseases using the CTD database, IL6 (Interleukin 6) and PTGS2/COX2 were the top hits, with 108 and 105 MeSH terms respectively, followed by TGFβ1 (transforming growth factor, beta 1) with 76 MeSH terms, and NOS2 (nitric oxide synthase 2) and AGT (Angiotensinogen) with 69 and 64 MeSH terms, respectively. Common to most of these genes is their key role in inflammatory responses, which underlies a myriad of diseases. For example, IL6 is produced at the site of inflammation, playing a





**Fig. 3.** Genetic disorders associated with  $Ca^{2+}$  genes. **(A)** Sunburst diagram of the MeSH term hierarchy. Fields/terms with at least one  $Ca^{2+}$  gene are labeled red, and those without associations are labeled blue. The sector angles correspond to the number of children each term includes. Sub-terms are further away from the center than parent terms. **(B)** Bar chart of the MeSH parent terms (inner circle of panel A) and the cumulative associated  $Ca^{2+}$  genes. **(C)** Heat map of Pearson's correlation values of diseases and hierarchical clustering of groups. Each color shows clusters selected by consensus clustering of  $k=5$ . **(D)** Disease names analyzed for term frequency of each cluster, filtered for the top 10% of terms with the highest variance

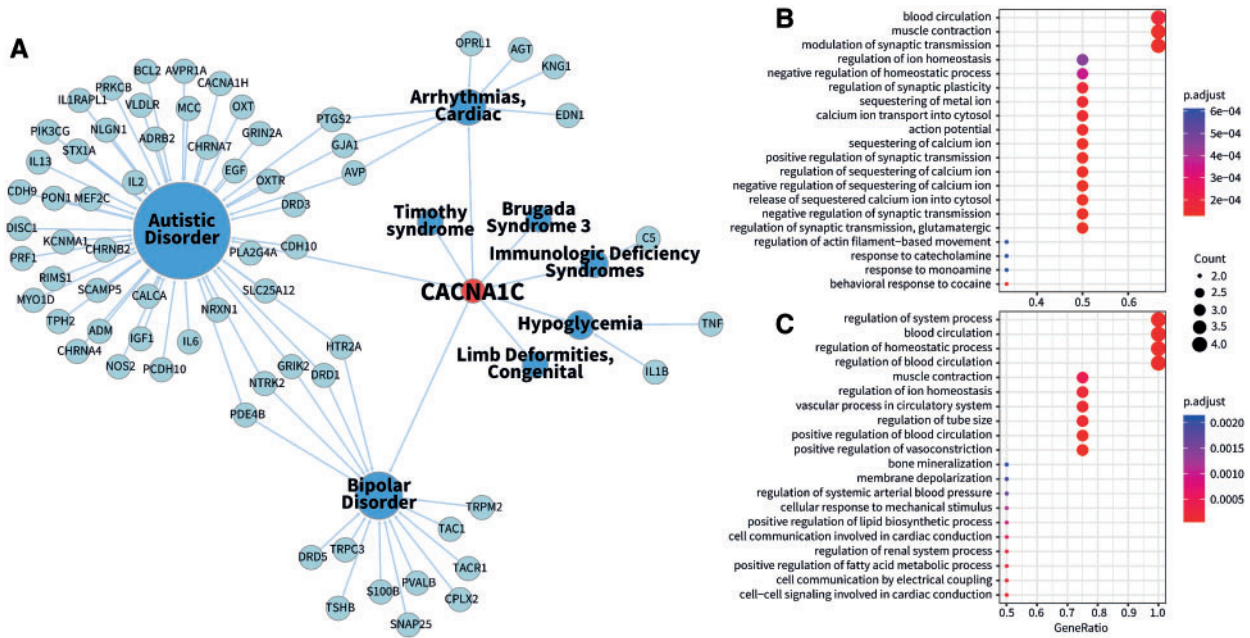
key role in the acute phase response (Tanaka *et al.*, 2014), PTGS2/COX2 in the production of inflammatory prostaglandins from arachidonic acid (Alhouayek and Muccioli, 2014). Additionally NOS2, as well as being a messenger molecule with diverse functions, mediates cysteine S-nitrosylation of PTGS2/COX2 and enhances the synthesis of IL6. Interestingly, NOS2 transcription is modulated by the  $Ca^{2+}$  dependent transcription factor NF- $\kappa$ B (nuclear Factor kappa-light-chain-enhancer of activated B cells) (Zhou *et al.*, 2011). TGFB1, as well as functioning as a proinflammatory cytokine, plays an important role in bone remodeling, a pivotal role in tumor suppression (and paradoxically tumor promotion) and can induce epithelial-to-mesenchymal transition (Hwangbo *et al.*, 2016). Whereas AGT is an essential component of the renin-angiotensin system, being a potent regulator of blood pressure, body fluid and electrolyte homeostasis.

We also sought to determine whether diseases could be grouped together in an unbiased manner with regards to their association with  $Ca^{2+}$  signaling genes to reveal relationships between diseases with no apparent connections through the clustering of genes involved in  $Ca^{2+}$  signaling. Hence, diseases were clustered based on their number of shared  $Ca^{2+}$  genes. Generating a correlation matrix based on the entire  $Ca^{2+}$  genome yielded five major groups of clustered diseases (Fig. 3C). We hypothesized that the clusters corresponded to diseases with shared etiologies or mechanisms of action for the pathology itself. Interesting relations between diseases occurred when the disease names were analyzed by cluster (Fig. 3D). For example, Cluster 1 was rather homogeneous because it mainly contained developmental disorders, such as muscular dystrophy, hereditary nephritis (Alport Syndrome), and Alveolar capillary dysplasia, whereas Cluster 2 was more heterogeneous and included diverse diseases, such as multiple sclerosis, diabetes mellitus, chronic kidney failure and pulmonary fibrosis. Cluster 5 contained mostly cancer related diseases, such as skin neoplasms, B-cell lymphoma and breast carcinoma. Additionally, Cluster 5 also contained asthma and amyotrophic lateral sclerosis.

### 3.4 CACNA1C-associated gene network

Finally, we wanted to perform an analysis of one  $Ca^{2+}$  gene that is known to be involved in several diseases, preferably in different organs. Thus, we searched our database for genetic disorders associated with CACNA1C, which encodes the alpha-1 subunit of the voltage-dependent  $Ca^{2+}$  channel  $Ca_v1.2$ . Eight genetic disorders with mutations in the CACNA1C gene were found: cardiac arrhythmia, autistic disorder, bipolar disorder, brugada syndrome 3, hypertension, hypoglycemia, immunologic deficiency syndromes, congenital limb deformities and Timothy syndrome. These diseases are mainly caused by mutations in CACNA1C that occur during embryonic or fetal development, although the effects may not be observed until later in life.

Plotting a network map that was expanded one degree revealed 67 genes linked to CACNA1C-related diseases (Fig. 4). Of the 8 associated diseases, four (timothy syndrome, Brugada Syndrome 3, Limb Deformities, congenital) have CACNA1C as their sole  $Ca^{2+}$ -related gene involved, whilst the other remaining diseases had one or more  $Ca^{2+}$ -related genes linked with the disease. The two CACNA1C-associated neurological disease families, autistic disorder and bipolar disorder, of which both had the highest numbers of associated calcium-related genes, shared the following genes: PDE4B, NTRK2, GRIK2, DRD1 and HTR2A. Enrichment analysis of the GO annotations linked to these particular genes revealed that they are involved in a range of biological processes including the modulation of synaptic transmission, synaptic plasticity, blood circulation, muscle contraction and regulation of actin filament based movements (Fig. 4B). Comparing shared calcium related genes between cardiac arrhythmias and autistic disorder (PTGS2, GJA1 and AVP), the enrichment analysis of the GO ontology revealed biological processes ranging from the regulation of system and homeostatic processes, vasoconstriction, renal system processes, cell communication by electrical coupling and regulation of lipid biosynthesis (Fig. 4C).



**Fig. 4.** CACNA1C disease network with associated  $\text{Ca}^{2+}$  genes. **(A)** Network of disease categories (blue) linked to CACNA1C (red) expanded one degree, which indicated 67 associated  $\text{Ca}^{2+}$  genes (green). In **(B)** and **(C)**, dot plot displaying the top 20 GO terms from the GO-enrichment analysis of the genes common to: **(B)** Autistic Disorder and Bipolar Disorder, **(C)** Autistic Disorder and Cardiac Arrhythmias. Color bar code display the adjusted  $P$ -values for multiple comparisons. Dot sizes display number of associated genes

The analysis of the interrelationships among the  $\text{Ca}^{2+}$  signaling genes and their associated diseases will not only increase our general understanding of how genetic disorders are connected and develop but also open up new approaches for treating these common diseases.

#### 4 Discussion

Our results demonstrate the key role of  $\text{Ca}^{2+}$  in regulating important cell processes in virtually all of the different cell types of the human body. First, all of the genes involved in intracellular  $\text{Ca}^{2+}$  signaling in human cells were mapped. Second, human genetic disorders with reported mutations in the genes associated with  $\text{Ca}^{2+}$  signaling were determined. Third, an automatically updated online research tool (<http://cagedb.uhlenlab.org>) for identifying genes and diseases linked to  $\text{Ca}^{2+}$  signaling was established.

Central to this study are the GO terms, which facilitate the identification of genes related to  $\text{Ca}^{2+}$  signaling. New GO terms are constantly established as a result of novel scientific discoveries of previously annotated genes (Blake, 2013). By applying text-mining algorithms to scientific publications, new GO term annotations can be determined (Bayes et al., 2014). Here, we used only the assigned annotations. A quarter of all of the annotations of our genes were based on experimental data, which indicates reliability (Skunca et al., 2012). The set of GO terms used in the online research tool implemented here is manually updated and currently consists of 241 terms.

Our unbiased clustering of diseases based on genes involved in  $\text{Ca}^{2+}$  signaling, which is presented in Figure 3C, D, is an example of how novel hypotheses can be generated using this online research tool. The results of this analysis can be investigated further by determining the overlapping genes between diseases with no apparent connections. For example, our analysis revealed connections

between amyotrophic lateral sclerosis and various cancers. Such information can, for instance, be used to test if a successful treatment for a certain disease is applicable for another disease because of their overlapping  $\text{Ca}^{2+}$  signaling genes. Moreover, the gene network for CACNA1C presented in Figure 4 is another example of how this tool can be used. This network showed that the genes PTGS2, GJA1 and AVP as well as CACNA1C were all linked to the two different organ disease groups autistic disorder and cardiac arrhythmias. The PTGS2 gene is linked to the greatest number of diseases. GJA1 (Gap junction alpha-1) is also known as Cx43 (connexin 43), a protein that enables the conduction of  $\text{Ca}^{2+}$  between cells and the extracellular space via gap junctions and hemichannels. AVP (arginine vasopressin) is a posterior pituitary hormone that has antidiuretic effects on the kidney and can also contract smooth muscles during parturition and lactation. AVP is also involved in cognition, tolerance, adaptation and complex sexual and maternal behavior. Additionally, the network analysis of CACNA1C related diseases showed that the shared genes between the two psychiatric disorders were mainly involved in synaptic transmission, blood circulation and muscle contraction. Interestingly, evaluation of the excitability and viability of central motor pathways of the human motor cortex by transcranial magnetic stimulation has been widely used for the investigation of a variety of neurologic and psychiatric disorders (Chroni et al., 2002). Additionally, changes observed in autism include alterations in the signaling pathways mediating neurovascular coupling as a result of an increase in synaptic inhibition (Reynell and Harris, 2013). Cerebral blood flow and cerebral metabolic rate are closely correlated and are conceptualized as proxies of synaptic transmission with a number of developmental, degenerative, neoplastic and ischemic processes associated with an uncoupling of blood flow and metabolism (Ota et al., 2014).

During the identification of the  $\text{Ca}^{2+}$  genes and their related diseases for this study, the strength of the different associations varied considerably. The strength (or evidence) of a certain gene-disease

association was based on a number of methods and standards for measuring, validating and interpreting genetic associations (Khoury *et al.*, 2007). Additionally, the selection criteria used to identify the genes involved in Ca<sup>2+</sup> signaling were not always entirely reliable because the Gene Ontology database does not contain this specific category. However, we decided to risk the inclusion of false positives, i.e. gene-disease associations that are not actually related to Ca<sup>2+</sup> signaling, rather than miss true gene-disease associations.

Annotations are either automatically generated or manually added to databases by curators. These processes can certainly be further improved, which would result in more specific and up-to-date annotations (du Plessis *et al.*, 2011). Currently, scientists must report their own findings to be absolutely sure that the results are annotated accurately in the associated databases. The purpose of this work was to survey available databases to identify genes related to Ca<sup>2+</sup> signaling and their associated diseases. As a consequence, a number of Ca<sup>2+</sup> gene-disease associations were most likely not included because they have not yet been annotated. In the future, further improvements to machine learning will increase the reliability of the annotation process.

In summary, we have mapped 1805 genes and 1470 human genetic disorders that are involved in Ca<sup>2+</sup> signaling. Our online research tool (<http://cagedb.uhlenlab.org>) is a dynamic database that is updated weekly with new genes and genetic disorders associated with Ca<sup>2+</sup> signaling. This online database will be a valuable resource that should help facilitate a better understanding of the network interactions between Ca<sup>2+</sup> signaling genes and diseases. Our study confirms the important role of Ca<sup>2+</sup> in living cells and shows that perturbed Ca<sup>2+</sup> handling can result in a large number of diverse human diseases.

## Acknowledgements

We thank Dr. Sten Linnarsson and Dr. Peter Lönnerberg of the Karolinska Institutet for their helpful discussions.

## Funding

This work was supported by the Swedish Research Council (grants 2009-3364, 2010-4392 and 2013-3189 to PU), the Swedish Strategic Foundation (MultiBIO 2010 to PU), the Swedish Cancer Society (grant CAN 2013-802 and CAN 2016-801 to PU), the Linnaeus Center in Developmental Biology for Regenerative Medicine (DBRM to PU), the Sigrid Jusélius Foundation (to LL), the Knut and Alice Wallenberg Foundation (grants CLICK and Research Fellow to PU) and the Swedish Brain Foundation (grant FO2014-0220 and FO2015-0074 to PU).

*Conflict of Interest:* none declared.

## References

Ashouyeh, M. and Muccioli, G.G. (2014) COX-2-derived endocannabinoid metabolites as novel inflammatory mediators. *Trends Pharmacol. Sci.*, **35**, 284–292.

Ashburner, M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

Bayes, A. *et al.* (2014) Human post-mortem synapse proteome integrity screening for proteomic studies of postsynaptic complexes. *Mol. Brain*, **7**, 88.

Berridge, M.J. (2012) Calcium signalling remodelling and disease. *Biochem. Soc. Trans.*, **40**, 297–309.

Berridge, M.J. *et al.* (2003) Calcium signalling: dynamics, homeostasis and remodelling. *Nat. Rev. Mol. Cell Biol.*, **4**, 517–529.

Berridge, M.J. *et al.* (2000) The versatility and universality of calcium signalling. *Nat. Rev. Mol. Cell Biol.*, **1**, 11–21.

Bezprozvanny, I. and Hayden, M.R. (2004) Deranged neuronal calcium signaling and Huntington disease. *Biochem. Biophys. Res. Commun.*, **322**, 1310–1317.

Bezprozvanny, I. and Mattson, M.P. (2008) Neuronal calcium mishandling and the pathogenesis of Alzheimer's disease. *Trends Neurosci.*, **31**, 454–463.

Bidaud, I. *et al.* (2006) Voltage-gated calcium channels in genetic diseases. *Biochim. Biophys. Acta*, **1763**, 1169–1174.

Blake, J.A. (2013) Ten quick tips for using the gene ontology. *PLoS Comput. Biol.*, **9**, e1003343.

Bootman, M.D. *et al.* (2009) An update on nuclear calcium signalling. *J. Cell Sci.*, **122**, 2337–2350.

Burashnikov, E. *et al.* (2010) Mutations in the cardiac L-type calcium channel associated with inherited J-wave syndromes and sudden cardiac death. *Heart Rhythm.*, **7**, 1872–1882.

Chakraborti, S. *et al.* (2007) Calcium signaling phenomena in heart diseases: a perspective. *Mol. Cell. Biochem.*, **298**, 1–40.

Chroni, E. *et al.* (2002) Effect of exercise on motor evoked potentials elicited by transcranial magnetic stimulation in psychiatric patients. *J. Clin. Neurophysiol.*, **19**, 240–244.

Clapham, D.E. (2007) Calcium signaling. *Cell*, **131**, 1047–1058.

Davis, A.P. *et al.* (2015) The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. *Nucleic Acids Res.*, **43**, D914–D920.

Davis, A.P. *et al.* (2012) MEDIC: a practical disease vocabulary used at the Comparative Toxicogenomics Database. *Database (Oxford)*, **2012**, bar065.

Delmas, P. *et al.* (2004) Polycystins, calcium signaling, and human diseases. *Biochem. Biophys. Res. Commun.*, **322**, 1374–1383.

Du Plessis, L. *et al.* (2011) The what, where, how and why of gene ontology—a primer for bioinformaticians. *Brief. Bioinform.*, **12**, 723–735.

Furukawa, K. *et al.* (2003) Alteration in calcium channel properties is responsible for the neurotoxic action of a familial frontotemporal dementia tau mutation. *J. Neurochem.*, **87**, 427–436.

Gene Ontology, C. (2008) The Gene Ontology project in 2008. *Nucleic Acids Res.*, **36**, D440–D444.

Giacomello, M. *et al.* (2013) Plasma membrane calcium ATPases and related disorders. *Int. J. Biochem. Cell Biol.*, **45**, 753–762.

Hernandez-Ochoa, E.O. *et al.* (2015) Critical role of intracellular RyR1 calcium release channels in skeletal muscle function and disease. *Front. Physiol.*, **6**, 420.

Hwangbo, C. *et al.* (2016) Syntenin regulates TGF-beta1-induced Smad activation and the epithelial-to-mesenchymal transition by inhibiting caveolin-mediated TGF-beta type I receptor internalization. *Oncogene*, **35**, 389–401.

Jackman, R.W. and Kandarian, S.C. (2004) The molecular basis of skeletal muscle atrophy. *Am. J. Physiol. Cell Physiol.*, **287**, C834–C843.

Khoury, M.J. *et al.* (2007) On the synthesis and interpretation of consistent but weak gene-disease associations in the era of genome-wide association studies. *Int. J. Epidemiol.*, **36**, 439–445.

Kristensson, K. *et al.* (1993) Scrapie prions alter receptor-mediated calcium responses in cultured cells. *Neurology*, **43**, 2335–2341.

Kuo, I.Y. and Ehrlich, B.E. (2012) Ion channels in renal disease. *Chem Rev.*, **112**, 6353–6372.

Malmersjö, S. *et al.* (2013) Neural progenitors organize in small-world networks to promote cell proliferation. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, E1524–E1532.

Martino, D. *et al.* (2009) Immunopathogenic mechanisms in tourette syndrome: a critical review. *Mov. Disord.*, **24**, 1267–1279.

Nedergaard, M. *et al.* (2010) Glial calcium and diseases of the nervous system. *Cell Calcium*, **47**, 140–149.

Osman, A.A. *et al.* (2003) Wolframin expression induces novel ion channel activity in endoplasmic reticulum membranes and increases intracellular calcium. *J. Biol. Chem.*, **278**, 52755–52762.

Ota, M. *et al.* (2014) Altered coupling of regional cerebral blood flow and brain temperature in schizophrenia compared with bipolar disorder and healthy subjects. *J. Cereb. Blood Flow Metab.*, **34**, 1868–1872.

Priori, S.G. *et al.* (2001) Mutations in the cardiac ryanodine receptor gene (hRyR2) underlie catecholaminergic polymorphic ventricular tachycardia. *Circulation*, **103**, 196–200.

- Rajakulendran,S. et al. (2012) Neuronal P/Q-type calcium channel dysfunction in inherited disorders of the CNS. *Nat. Rev. Neurol.*, **8**, 86–96.
- Ramasamy,I. (2008) Inherited disorders of calcium homeostasis. *Clin. Chim. Acta Int. J. Clin. Chem.*, **394**, 22–41.
- Reynell,C. and Harris,J.J. (2013) The BOLD signal and neurovascular coupling in autism. *Dev. Cogn. Neurosci.*, **6**, 72–79.
- Rowland,L.P. and Shneider,N.A. (2001) Amyotrophic lateral sclerosis. *N. Engl. J. Med.*, **344**, 1688–1700.
- Skunca,N. et al. (2012) Quality of computationally inferred gene ontology annotations. *PLoS Comput. Biol.*, **8**, e1002533.
- Smedler,E. et al. (2014) Network analysis of time-lapse microscopy recordings. *Front. Neural Circ.*, **8**, 111.
- Suzuki,R. and Shimodaira,H. (2006) Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, **22**, 1540–1542.
- Tanaka,T. et al. (2014) IL-6 in inflammation, immunity, and disease. *Cold Spring Harb. Perspect. Biol.*, **6**, a016295.
- Thomas,B. and Beal,M.F. (2007) Parkinson’s disease. *Hum. Mol. Genet.*, **16**, R183–R194.
- Triplett,D.A. (2000) Coagulation and bleeding disorders: review and update. *Clin. Chem.*, **46**, 1260–1269.
- Uhlen,P. et al. (2006) Gain-of-function/Noonan syndrome SHP-2/Ptrpn11 mutants enhance calcium oscillations and impair NFAT signaling. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 2160–2165.
- Uhlen,P. and Fritz,N. (2010) Biochemistry of calcium oscillations. *Biochem. Biophys. Res. Commun.*, **396**, 28–32.
- Wilkerson,M.D. and Hayes,D.N. (2010) ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*, **26**, 1572–1573.
- Zemaitaitis,M.O. et al. (2000) Transglutaminase-induced cross-linking of tau proteins in progressive supranuclear palsy. *J. Neuropathol. Exp. Neurol.*, **59**, 983–989.
- Zhivotovsky,B. and Orrenius,S. (2011) Calcium and cell death mechanisms: a perspective from the cell death community. *Cell Calcium*, **50**, 211–221.
- Zhou,J. et al. (2011) A20-binding inhibitor of NF-kappaB (ABIN1) controls Toll-like receptor-mediated CCAAT/enhancer-binding protein beta activation and protects from inflammatory disease. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, E998–1006.