Genetic Analysis of Rare Human Variants of Regulators of G Protein Signaling Proteins and Their Role in Human Physiology and Disease

Katherine E. Squires, Carolina Montañez-Miranda, Rushika R. Pandya, Matthew P. Torres, and John R. Hepler

Abstract

Regulators of G protein signaling (RGS) proteins modulate the physiologic actions of many neurotransmitters, hormones, and other signaling molecules. Human RGS proteins comprise a family of 20 canonical proteins that bind directly to G protein–coupled receptors/G protein complexes to limit the lifetime of their signaling events, which regulate all aspects of cell and organ physiology. Genetic variations account for diverse human traits and individual predispositions to disease. RGS proteins contribute to many complex polygenic human traits and pathologies such as hypertension, atherosclerosis, schizophrenia, depression, addiction, cancers, and many others. Recent analysis indicates that most human diseases are

I. Introduction

A. G Protein Signaling

B. A (Very) Brief History of Regulators of G Protein Signaling

C. RGS Protein Rare Human Variants in Complex Diseases

II. RGS Proteins in Physiology and Human Disease

A. The R4 Family

1. R4 Family Overview

2. R4 Family Proteins in Human Physiology and Disease

B. The R7 Family

1. R7 Family Overview

2. R7 Family Proteins in Human Physiology and Disease

C. The R12 Family

1. R12 Family Overview

2. R12 Family Proteins in Human Physiology and Disease

D. The RZ Family

1. RZ Family Overview

2. RZ Family Proteins in Human Physiology and Disease

III. Analysis of Rare Human Variants of RGS Proteins

A. R4 Family: RGS4 Rare Variants

B. R7 Family: RGS9 Rare Variants

C. R12 Family: RGS10 Rare Variants

D. RZ Family: RGS17 Rare Variants

IV. Pharmacological Impact of Human Variants in RGS Proteins

Acknowledgments

References

Address correspondence to: Dr. John R. Hepler, Emory University School of Medicine, G206 Rollins Research Center, 1510 Clifton Road G205, Atlanta, GA 30329-3090. E-mail: jhepler@emory.edu

This work was supported, in whole or in part, by National Institutes of Health [Grants 5R01NS037112 and 1R21 NS087488, both awarded to J.R.H., and Grant R01-GM117400 to M.P.T.]. R.R.P. was supported by the GT Bioinformatics Faculty Research Award, and K.E.S. was supported by National Institutes of Health [Training Grant T32 GM008602].

https://doi.org/10.1124/pr.117.015354.

This article has supplemental material available at pharmrev.aspetjournals.org.
I. Introduction

A. G Protein Signaling

Cells communicate by releasing chemical messengers that dictate cell and organ physiology. These messengers include many natural ligands such as hormones, neurotransmitters, cytokines, nutrients, sensory input, and other natural molecules. Because most of these messengers cannot cross the outer cell (plasma) membrane, membrane-bound receptors and their cognate heterotrimeric G proteins (Goβγ) act as cellular transducers. At rest, G protein–coupled receptors (GPCRs) are in close proximity with an inactive Goβγ protein complex, with GDP bound to Go. Upon receptor activation by ligand binding, the chemical message is transduced across the plasma membrane through the receptor to promote the release of GDP from the associated Go. Due to the high concentration of GTP in the cytosol, Go quickly binds a GTP to form Go-GTP, which has reduced affinity for both receptor and Goβγ, and thus dissociates. The free Go-GTP and Goβγ are then able to relay the chemical message to an intracellular cascade of effectors and second messengers that dictate all aspects of cell and organ physiology. The duration of signaling by Go-GTP and Goβγ is governed by the intrinsic GTPase activity of the Go subunit (Gilman, 1987; Bourne et al., 1990; Simon et al., 1991; Hepler and Gilman, 1992; Hamm, 1998). Upon hydrolysis of the GTP to GDP, Go-GDP reassociates with Goβγ to terminate signaling. Under normal physiologic conditions, a Go acts as a molecular switch capable of turning itself off by GTP hydrolysis. However, researchers noted early on that the rate of GTP hydrolysis by purified Go proteins in vitro was much slower than that observed by Go in cells, hinting at a missing piece of the puzzle: cellular factor(s) that governs the off rate of G protein signaling (Wagner et al., 1988; Vuong and Chabre, 1990).

B. A (Very) Brief History of Regulators of G Protein Signaling

The missing piece was identified based on foundational observations dating back to the early 1980s. SST2, a yeast protein important for mating, was found to regulate sensitivity to pheromones (Chan and Otte, 1982). Later it was shown that SST2 does so via regulation of Dohlman et al. (1995) and physical association with (Dohlman et al., 1996) yeast Go proteins, which share homology to mammalian systems. These findings followed the discovery of a class of mammalian proteins that could enhance the GTPase activity of Ras and Ras-like small G proteins (Trahey and McCormick, 1987), speeding up the turnover of GTP and accordingly the termination of signaling [termed GTPase-accelerating proteins (GAPs)]. The postulated missing piece, a mammalian GAP for heterotrimeric Go proteins, was discovered soon thereafter as a family of novel proteins very similar to the yeast SST2 (De Vries et al., 1995; Druey et al., 1996; Koelle and Horvitz, 1996). These specialized Go GAPs were named regulators of G protein signaling (RGS), and thus the RGS field was born.

Since their initial discovery, 20 canonical RGS (Fig. 1) and 19 RGS-like proteins have been identified, and extensive characterization of these proteins has revealed multifunctional roles (Ross and Wilkie, 2000; Hollinger and Hepler, 2002; Willars, 2006; Evans et al., 2015). All canonical RGS proteins share a conserved, approximately 120 amino acid RGS domain, which binds active Go-GTP and catalyzes the transition state of GTP hydrolysis by Go, demonstrated by the fact that the RGS domains bind preferentially to Go-GDP activated with the transition state mimetic, AlF4− (Tesmer et al., 1997). Although RGS proteins are classified according to the presence of a RGS domain, we now understand that RGS proteins encompass a wide diversity of multidomain signaling and scaffolding proteins that are categorized by sequence similarity (Fig. 1). Furthermore, we have come to appreciate that regulation of G protein signaling is crucial for normal cellular function, and improperly regulated G protein signaling underlies many disease states (Gerber et al., 2016; Sjögren, 2017), signifying the potential for RGS proteins as therapeutic targets. In this review, we limit and focus our discussion to the 20 canonical RGS proteins. Due to their sequence heterogeneity, we have grouped the discussion and analysis of each RGS within its respective RGS subfamily.

ABBREVIATIONS: CADD, combined annotation-dependent depletion; CoF, change of function; DEP, disheveled EGL10-Pleckstrin; ERK, extracellular signal-regulated kinase; GAP, GTPase-accelerating protein; GGL, G protein γ subunit-like; GoF, gain of function; GPCR, G protein–coupled receptor; GPR, G protein regulatory; LoF, loss of function; MAP, modified alignment position; MTR, missense tolerance ratio; PTM, post-translational modification; RGS, regulators of G protein signaling; SNP, single nucleotide polymorphism; SNV, single nucleotide variant.
C. RGS Protein Rare Human Variants in Complex Diseases

Recent genetic analysis of human exomes indicates that an explosion of variants within protein-coding regions arose between 5000 and 10,000 years ago, leading to diverse human traits and a broad range of potential disease determinants (Fu et al., 2013). Population and clinical genome sequencing has generated new information regarding the etiology of disease, and the recent release of large-scale genome and exome sequencing data [Genome Aggregation Database, GnomAD, of the Broad Institute (Lek et al., 2016)] has allowed for the analysis and possible discovery of new disease-linked rare variants. Although some proteins are functionally well-positioned to participate in monogenic diseases (Yuan et al., 2015a; Ogden et al., 2017), others such as RGS proteins are likely involved in more complex pathways that contribute to disease progression (Bansal et al., 2007; Gerber et al., 2016; Xie et al., 2016). Whereas genetic deletion of the ubiquitously expressed heterotrimeric G proteins can result in severe defects or embryonic lethality (Offermanns et al., 1997, 1998; Yu et al., 1998; Wettischreuck and Offermanns, 2005; Okae and Ikawura, 2010; Plummer et al., 2012; Moon et al., 2014), genetic loss of their more discretely expressed modulators, the RGS proteins, are likely involved in more complex pathways that contribute to disease progression (Bansal et al., 2007; Gerber et al., 2016; Xie et al., 2016). Whereas genetic deletion of the ubiquitously expressed heterotrimeric G proteins can result in severe defects or embryonic lethality (Offermanns et al., 1997, 1998; Yu et al., 1998; Wettischreuck and Offermanns, 2005; Okae and Ikawura, 2010; Plummer et al., 2012; Moon et al., 2014), genetic loss of their more discretely expressed modulators, the RGS proteins, results (broadly speaking) in subtle and less detrimental phenotypes with only a few exceptions (see below). RGS proteins therefore participate in complicated multifactorial physiologic processes and disease states (Williams et al., 2004; Zhang and Mende, 2014; Evans et al., 2015; Ganss, 2015; Ahlers et al., 2016), and large-scale genomic association studies may not be able to detect disease-associated polymorphisms (Bomba et al., 2017). Due to the purifying drive of natural selection, very rare variants (1%–2% or less) in fact seem to play a far greater role as genetic determinants of disease compared with the more common and less deleterious polymorphisms (Nelson et al., 2012; Tennessen et al., 2012), an idea supported by the observed inverse relationship between minor allele frequency and disease risk (Park et al., 2011). The rarity of these variants means that most carriers are heterozygous, as the likelihood of inheriting the same single nucleotide variant (SNV) from each parent is incredibly low. However, both diseases and traits can arise from heterozygosity. For instance, a single variant allele can cause a dominant-negative phenotype (Kenakin and Miller, 2010), an intermediate phenotype (as compared with wild type or knockout) (Okamoto et al., 2017), mislocalization of proteins to cellular compartments leading to aberrant singling (Lo Bello et al., 2017), or a change of function (CoF) in other ways.

Given this context, in this work we describe a genomics approach to identify likely deleterious rare human variants (defined in this study as less than 2% prevalence) in functionally sensitive regions of RGS proteins, with an emphasis on the well-defined RGS domain. Due to the likely involvement of RGS proteins in complex disease states (Sjögren, 2017), these variants may contribute to a first hit, leaving the carrier more vulnerable to disease given a secondary insult (e.g., environmental, subsequent mutation of a protein in related signaling pathway, or otherwise). Furthermore, both a loss-of-function (LoF) as well as a gain-of-function (GoF) variant could equally disrupt cellular systems in delicate equilibrium. Thus, we suggest an unbiased approach that measures a CoF, which encompasses both variant forms. Finally, it is important to consider that, due to cultural practices, variants may be

![Phylogenetic tree of the human RGS protein family](image-url)
expressed more commonly or exclusively within a single ethnic group, and may give rise to unique human traits, rather than the more obvious and deleterious disease states.

In this review, we examine the 20 canonical human RGS proteins (Fig. 1) and their known links to disease or traits, and then present human variants for these gene exomes, extracted from the publically available GnomAD project of the Broad Institute (Lek et al., 2016). In our analysis, we use recently developed bioinformatics, proteomic and structural tools, including combined annotation-dependent depletion (CADD) (Kircher et al., 2014), missense tolerance ratio (MTR) (Traynelis et al., 2017), and post-translational modification (PTM) cluster analysis (Dewhurst et al., 2015; Torres et al., 2016).

With the availability of large-scale databases for human exome sequences comes the major challenge in medical genomics of determining the significance of (or lack thereof) any particular genetic variant. At present, variants of uncertain significance represent the vast majority of those described in genetic reports. MTR offers a new method for summarizing available human variation data within genes to capture population level genetic variation and measures the ratio of tolerance within exome sequences to genetic mutations (Traynelis et al., 2017). This analysis uses publicly available human mutation data to make predictions about domains and motifs that are resistant to, and atypically low in number of, missense mutations. From these ratios, we can predict which exome regions are likely to be functionally sensitive to mutation. Integrating MTR with other selected bioinformatic tools (e.g., CADD and PTM cluster analysis) for any particular exome sequence provides a way of predicting pathogenic missense variants from background missense variation in disease genes. CADD is a complementary bioinformatic measure of mutation severity, but takes into account multiple measures of deleteriousness. Unlike MTR, CADD analysis tells us not where, but how a mutation might affect protein function. CADD independently integrates various diverse annotations into a single measure (C score) for each variant. The score takes into account measures of sequence conservation and amino acid side chain properties and prioritizes functional, deleterious, and pathogenic variants across many functional categories, effect sizes, and genetic architectures to provide researchers with a valuable tool for selecting variants for study. PTMs play critical roles in regulating and determining protein function, the disruption of which can cause disease (Jensen et al., 2002; Hornbeck et al., 2012; Lothrop et al., 2013; Dewhurst et al., 2015; Torres et al., 2016). Inherently, PTMs alter the structure of proteins and therefore have the potential to alter their function as well. In addition, PTMs can have pronounced effects on protein–protein interactions, serving in many cases as handles for PTM-specific binding domains (Walsh et al., 2005). A handful of experimentally verified PTMs is reported in human RGS proteins (Alqinyah and Hooks, 2018), including phosphorylation, ubiquitination, acetylation, methylation, and palmitoylation. The function of most PTMs (in RGS or other proteins) remains undetermined due in large part to the time and challenge requirements of conducting biochemical experiments as well as a dearth of suitable methods for functional prioritization of existing knowledge.

Due to their important roles in protein structure and function, we identified and reported a comprehensive list of experimentally verified and publically curated PTMs found on human RGS proteins (Li et al., 2014; Huang et al., 2016b) and mapped them with respect to each RGS protein. For this study, we use a recently described protein family-specific PTM alignment analysis method that has proven utility for revealing functional PTM hotspots in protein families (Dewhurst et al., 2015; Torres et al., 2016), and we report those PTMs in RGS proteins that overlap with human variants. Using all of this information (human variants, CADD, and PTMs) in combination, we have prioritized a narrow list of select variants that we predict will disrupt human RGS protein function. As a proof-of-principle to validate the approach, we test one of these selected variants to demonstrate a profound CoF, and highlight others for future study. Although we focus in this work on the RGS domain, we recognize that other RGS protein regions and domains are also essential for RGS protein function. As such, the comprehensive dataset examining these measures for the entire exome sequence for each canonical RGS protein is also provided for any investigator to use (Supplemental Material). We reason that if rare variants occur in functionally sensitive regions of RGS proteins (e.g., the RGS domain) such that they confer a profound CoF phenotype, then these variants likely make important (and previously unappreciated) contributions to complex human disease states and/or unique human traits. As such, we believe that computationally identified rare variants, combined with experiments that validate a CoF for such variants, can provide a deeper understanding of the etiology of complex disease states and the evolution of human traits, while also providing investigators and clinicians a valuable wealth of information toward personalized medicines.

II. RGS Proteins in Physiology and Human Disease

A. The R4 Family

1. R4 Family Overview. Members of the R4 subfamily compose the largest and best characterized of the RGS proteins due to their early discovery and simplicity in structure. R4 family RGS proteins include the following: RGS1, RGS2, RGS3, RGS4, RGS5, RGS8, RGS13, RGS16, RGS18, and RGS21 (Fig. 1). All R4 family members exhibit capacity to bind and act as GAPs for both Ga10 and Gaq (Hollinger and Hepler,
2. R4 Family Proteins in Human Physiology and Disease. Although not initially recognized as a RGS protein, RGS1 was cloned from activated B lymphocytes as an early activation gene, and designated BL34 (Hong et al., 1993) shortly before the RGS domain was characterized. Since its discovery, much has been learned about the role of RGS1 in immune physiology and pathology (Xie et al., 2016). In B and T lymphocytes, RGS1 controls Gαi2-mediated chemotaxis and migration (Hwang et al., 2010; Gibbons et al., 2011). RGS1 knockout mice exhibited abnormal B cell migration, exaggerated germinal center formation, and atypical spleens (Moratz et al., 2004). Unsurprisingly, RGS1 has been linked with multiple T- and B-cell–related diseases such as inflammatory bowel disease, multiple sclerosis (Johnson et al., 2010), type 1 diabetes (Smyth et al., 2008), and celiac disease (Hunt et al., 2008; Smyth et al., 2008). RGS1 also has been implicated in atherosclerosis, and is overregulated in atherosclerotic plaques and aortic aneurysms, where it regulates monocyte and macrophage chemotaxis (Patel et al., 2015). A recent report demonstrated that elevated RGS1 expression in diffuse large B-cell lymphoma was associated with poor overall survival (Carreras et al., 2017). In this light, LoF variants may play a role in RGS1-mediated immune disorders, or susceptibility to these disorders.

RGS2 is a widely expressed (Kehrl and Sinnarajah, 2002) immediate early gene that is induced by various stimuli (Song et al., 2001), although it is best understood for its roles in the vasculature and the brain. In the hippocampus, a brain region related to learning and memory, high frequency stimulation (a method used to produce synaptic plasticity and long-term potentiation) induces RGS2 expression (Ingi et al., 1998), and RGS2 knockout mice exhibit impaired basal neuronal activity (Oliveira-Dos-Santos et al., 2000), supporting a role for RGS2 in synaptic plasticity and memory (Gerber et al., 2016). Due to its widespread expression throughout the brain, RGS2 has been linked to many neurologic diseases and affective disorders, including anxiety, post-traumatic stress disorder, and suicide (Amstadter et al., 2009a,b; Koenen et al., 2009; Lifschytz et al., 2012; Okimoto et al., 2012; Hohoff et al., 2015). In particular, a variant in the 3′ untranslated region of RGS2, rs4606, is associated with reduced mRNA expression of RGS2 (Semplicini et al., 2006), and is linked to anxiety (Leygraf et al., 2006; Koenen et al., 2009) and suicide (Cui et al., 2008; Amstadter et al., 2009b). In the periphery, RGS2 is expressed in vascular smooth muscle cells, where it is regulated by nitric oxide and controls vascular relaxation (Tang et al., 2003; Sun et al., 2005). RGS2 knockout mice are hypertensive (Heximer et al., 2003), and hypertensive human patients have reduced RGS2 mRNA compared with controls (Semplicini et al., 2006). Furthermore, hypertensive patients were more likely to have a SNV in the 3′ untranslated region (rs4606), which correlated with RGS2 expression (Semplicini et al., 2006). In a Japanese cohort, several N-terminal coding mutations (Q2L, M5V, and R44H) were associated with hypertension (Yang et al., 2005). In a subsequent study, it was found that the Q2L mutation destabilized RGS2 protein (but was reversed by proteosomal inhibition), whereas Q2R (another SNV found in both cases and controls) did not (Bodenstein et al., 2007). A more recent study demonstrated that four human variants, including Q2L, R188H, R44H, and D40Y, showed reduced capacity to inhibit Ca2+ release by the angiotensin II receptor (Phan et al., 2017). For further review of RGS2 function in physiology, see Bansal et al. (2007). Altogether, RGS2 expression has been found to regulate multiple aspects of normal and pathophysiology, and LoF variants would likely generate similar phenotypes to loss of protein. Finally, RGS2 also has been shown to negatively regulate protein translation by binding directly to the eukaryotic initiation factor 2B (eIF2B) subunit via a 37–amino-acid segment within the RGS domain (Nguyen et al., 2009), suggesting variants in this region could impact protein translation.

In contrast to RGS2, less is known about RGS3. RGS3 is found in the heart (Zhang et al., 1998), and mouse overexpression studies have shown that it protects against cardiac hypertrophy (Liu et al., 2014) and also regulates the survival/differentiation responses in multiple cell types (Nishiuma et al., 2009; Qiu et al., 2010). In humans, the longest splice variant (and canonical PDZ-containing isoform) RGS3-1 may promote epithelial–mesenchymal
<table>
<thead>
<tr>
<th>RGS Family</th>
<th>RGS</th>
<th>Major Tissue Distribution</th>
<th>Function</th>
<th>Knockout Phenotype</th>
<th>Proposed Disease Links</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>R4 subfamily</td>
<td>RGS1</td>
<td>Immune system (B and T lymphocytes, monocytes, natural killer cells, etc.)</td>
<td>Suppress chemotaxis</td>
<td>Abnormal B cell migration, exaggerated germinal center formation, and atypical spleens</td>
<td>Diseases of the immune system (inflammatory diseases, B cell lymphomas, etc.)</td>
<td>Moratz et al., 2004; Hunt et al., 2008; Smyth et al., 2008; Johnson et al., 2010; Patel et al., 2015</td>
</tr>
<tr>
<td></td>
<td>RGS2</td>
<td>Cardiovascular system and brain</td>
<td>Vascular relaxation</td>
<td>Impaired basal neuronal activity and hypertension</td>
<td>Anxiety, post-traumatic stress disorder, suicide, hypertension</td>
<td>Oliveira-Dos-Santos et al., 2000; Heximer et al., 2003; Yang et al., 2005; Leygraf et al., 2006; Semplicini et al., 2006; Amstadter et al., 2009a,b; Koenen et al., 2009; Lifschytz et al., 2012; Okimoto et al., 2012; Hohoff et al., 2015</td>
</tr>
<tr>
<td>RGS3</td>
<td>Heart</td>
<td></td>
<td>Still under investigation</td>
<td>Overexpression protects against cardiac hypertrophy</td>
<td>Possible link to cancers</td>
<td>Ooe et al., 2007; Liu et al., 2014; Chen et al., 2004; Williams et al., 2004; Prasad et al., 2005; Emilsson et al., 2006; Guo et al., 2006; Campbell et al., 2008a; So et al., 2008; Ding and Hegde, 2009; Ho et al., 2010; Felkin et al., 2011; Lee et al., 2012; Jaba et al., 2013; Opel et al., 2015</td>
</tr>
<tr>
<td>RGS4</td>
<td>Heart and brain</td>
<td>Respond to cardiovascular stress and modulate neuronal signaling</td>
<td></td>
<td></td>
<td>Schizophrenia, alcoholism</td>
<td>Owen et al., 2001; Chowdari et al., 2002, 2008; Mittmann et al., 2002; Chen et al., 2004; Williams et al., 2004; Prasad et al., 2005; Emilsson et al., 2006; Guo et al., 2006; Campbell et al., 2008a; So et al., 2008; Ding and Hegde, 2009; Ho et al., 2010; Felkin et al., 2011; Lee et al., 2012; Jaba et al., 2013; Opel et al., 2015</td>
</tr>
<tr>
<td>RGS5</td>
<td>Vasculature</td>
<td>Control vascular remodeling</td>
<td>Reduced blood pressure, improved immune-mediated destruction of solid tumors and survival</td>
<td>Hypertension, atherosclerosis, cancer angiogenesis</td>
<td>Chang et al., 2007; Cho et al., 2008; Hamzah et al., 2008; Nisancioglu et al., 2008; Holobotovskyy et al., 2013; Cheng et al., 2015; Daniel et al., 2016</td>
<td></td>
</tr>
<tr>
<td>RGS8</td>
<td>Brain (cerebellum)</td>
<td>Still under investigation</td>
<td>None noted</td>
<td>Possible link to seizures</td>
<td>Gold et al., 2002</td>
<td></td>
</tr>
<tr>
<td>RGS13</td>
<td>Immune system (B and T lymphocytes, mast cells, etc.)</td>
<td>Suppress immune response</td>
<td>Enhanced B cell responses and anaphylaxis</td>
<td>Leukemia/Lymphoma, asthma</td>
<td>Bansal et al., 2008a,b; Pise-Masison et al., 2009; Hwang et al., 2013; Sethakorn and Dulin, 2013; Raeder et al., 2015</td>
<td></td>
</tr>
<tr>
<td>RGS16</td>
<td>Retina, heart, brain, liver, immune system</td>
<td>Regulate inflammatory response and circadian rhythms</td>
<td>Enhanced inflammatory response in the lung and dysregulated circadian signaling</td>
<td>Circadian clocks/ morning wakefulness</td>
<td>Dæ et al., 2011; Hayasaaka et al., 2011; Shankar et al., 2012; Hu et al., 2016; Jones et al., 2016</td>
<td></td>
</tr>
<tr>
<td>RGS18</td>
<td>Platelets</td>
<td>Regulation of hemostasis</td>
<td>Reduced bleeding and hyper-responsive platelet activation</td>
<td>Aspirin resistance, amyotrophic lateral sclerosis, suicide</td>
<td>Schosser et al., 2011; Haggmark et al., 2014; Mao et al., 2014; Alshbool et al., 2015</td>
<td></td>
</tr>
<tr>
<td>RGS21</td>
<td>Bitter and sweet taste cells</td>
<td>Gustatory sensation</td>
<td>Unknown</td>
<td>Still under investigation</td>
<td>Sharma et al., 2016</td>
<td></td>
</tr>
<tr>
<td>R7 subfamily</td>
<td>RGS6</td>
<td>Brain and heart</td>
<td>Regulate drug reward, anxiety, and depression</td>
<td>Protection from alcohol-seeking behaviors, anxiety and depression-like behavior; enhanced dopamine neurodegeneration</td>
<td>Alcoholism, anxiety and depression, Parkinson’s disease, Alzheimer’s disease, and schizophrenia</td>
<td>Bifsha et al., 2014; Schizophrenia Working Group of the Psychiatric Genomics, 2014; Stewart et al., 2014; Moon et al., 2015; Stewart et al., 2015; Ahlers et al., 2016</td>
</tr>
<tr>
<td></td>
<td>RGS7</td>
<td>Brain</td>
<td>Regulates anxiety and drug reward</td>
<td>Enhanced reward, withdrawal and analgesia, and delayed tolerance to morphine</td>
<td>Panic disorder and multiple sclerosis</td>
<td>Crowe et al., 2001; Gelertner et al., 2001; Hohoff et al., 2009; McCauley et al., 2009</td>
</tr>
<tr>
<td></td>
<td>RGS9</td>
<td>Retina; brain striatum</td>
<td>Regulation of photoreceptor and striatal signaling</td>
<td>Disrupted striatal dopamine signaling related to</td>
<td>Retinal dysfunction</td>
<td>Nishiguchi et al., 2004; Cheng et al., 2007; Hartong et al., 2007; Stockman et al., 2008; Michaelides et al., 2010</td>
</tr>
</tbody>
</table>
transition via WNT signaling (Shi et al., 2012), and have a role in docetaxel-resistant breast cancer (Ooe et al., 2007). Other reports link RGS3 expression to various types of cancers, mediated by microRNAs, which bind the 3' untranslated region to regulate RGS3 expression (Chen et al., 2016; Wang et al., 2017a).

RGS4 is selectively expressed in the heart and the brain (Zhang et al., 1998; Ingi and Aoki, 2002), where its expression and protein stability is tightly controlled by its N terminus (Bodenstein et al., 2007; Bastin et al., 2012). Although RGS4 protein in the heart is found at very low levels under normal cardiac physiology (Stewart et al., 2012), it can be upregulated during pathophysiology and cardiac remodeling (Owen et al., 2001; Mittmann et al., 2002; Felkin et al., 2011; Lee et al., 2012; Jaba et al., 2013). Additionally, RGS4 knockout mice are susceptible to atrial fibrillation (Opel et al., 2015). A great deal is known about RGS4 in the brain. First, RGS4 mRNA is decreased in the frontal cortex of schizophrenic patients, whereas other R4 family transcripts were not (Mirnics et al., 2001). Furthermore, several noncoding SNVs [single nucleotide polymorphism (SNP)1, SNP4, SNP7, and SNP18] are found to be associated with RGS4 expression and schizophrenia diagnosis in multiple populations (Chowdari et al., 2002, 2008; Chen et al., 2004; Williams et al., 2004; Prasad et al., 2005; Guo et al., 2006; Campbell et al., 2008a; So et al., 2008). Aside from schizophrenia, RGS4 mRNA levels and SNVs have been linked with Alzheimer's disease (Emilsson et al., 2006) and alcoholism (Ho et al., 2010). RGS4 is associated with Parkinson's disease development, via decreased regulation of dopamine receptors and/or muscarinic acetylcholine receptors (Ding et al., 2006; Lerner and Kreitzer, 2012; Min et al., 2012), although results are varied (Ashrafi et al., 2017). In dopamine-depleted mice, a mouse model for Parkinson's disease, RGS4 is upregulated and inhibits M4 muscarinic autoreceptors, an effect that is mimicked by RGS4 infusion onto untreated cells (Ding et al., 2006). Remarkably, in that study, infusion of a RGS4 construct that lacks the N terminus (and is less potent at M4 receptors) reversed native RGS4-mediated attenuation of M4 signaling in dopamine-depleted mice. That is to say, this dysfunctional RGS4 demonstrated a dominant-negative effect on RGS4-mediated cell signaling, suggesting that fully functional knockouts may arise from only one LoF allele. This supports the prediction that, whereas rare variants overwhelmingly occur in a heterozygous manner (as discussed above), diseases and traits can still arise from one minor allele, in this case as a dominant-negative effect.

RGS5 is found broadly in tissues, including but not limited to the cardiovascular system and brain (Seki et al., 1998). Although there have been several genome-wide association studies suggesting a link between

---

**TABLE 1—Continued**

<table>
<thead>
<tr>
<th>RGS Family</th>
<th>RGS Major Tissue Distribution</th>
<th>Function</th>
<th>Knockout Phenotype</th>
<th>Proposed Disease Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>R12 subfamily</td>
<td>RGS11 retina</td>
<td>Regulation of photoreceptor signaling and neurodegeneration</td>
<td>Growth retardation and increased bone mass</td>
<td>Parkinson's disease, schizophrenia and chemoresistant ovarian cancer</td>
</tr>
<tr>
<td></td>
<td>RGS12 brain and immune system</td>
<td>Suppression of neuroinflammation and neurodegeneration</td>
<td>Growth retardation and increased bone mass</td>
<td>Parkinson's disease, schizophrenia and chemoresistant ovarian cancer</td>
</tr>
<tr>
<td></td>
<td>RGS13 brain, lung, testis, heart, bone, and spleen</td>
<td>Regulation of osteoclast differentiation and bone remodeling</td>
<td>Growth retardation and increased bone mass</td>
<td>Growth retardation and increased bone mass</td>
</tr>
<tr>
<td></td>
<td>RGS14 brain, spleen, and heart</td>
<td>Regulation of blood calcium homeostasis, long-term potentiation regulation, and regulation of drug dependence</td>
<td>Growth retardation and increased bone mass</td>
<td>Growth retardation and increased bone mass</td>
</tr>
<tr>
<td></td>
<td>RGS17 brain</td>
<td>Regulation of drug dependence and regulation of cell proliferation</td>
<td>Growth retardation and increased bone mass</td>
<td>Growth retardation and increased bone mass</td>
</tr>
<tr>
<td></td>
<td>RGS18 brain, lung, and liver</td>
<td>Potential links to cardiac development and regulation of cell proliferation</td>
<td>Growth retardation and increased bone mass</td>
<td>Growth retardation and increased bone mass</td>
</tr>
<tr>
<td></td>
<td>RGS19 brain, eye</td>
<td>Potential links to cardiac development and regulation of cell proliferation</td>
<td>Growth retardation and increased bone mass</td>
<td>Growth retardation and increased bone mass</td>
</tr>
<tr>
<td></td>
<td>RGS20 brain and eye</td>
<td>Still under investigation</td>
<td>Growth retardation and increased bone mass</td>
<td>Growth retardation and increased bone mass</td>
</tr>
</tbody>
</table>
RGS5 variants and neurologic disorders, such as schizophrenia and bipolar disorder (Campbell et al., 2008b; Smith et al., 2009), its role is best defined in the vasculature. In this work, RGS5 is expressed in arterial smooth muscle cells, and its expression is correlated with protection from hypertension and atherosclerosis (Holobotovskyy et al., 2013; Cheng et al., 2015; Daniel et al., 2016). Specifically, RGS5 is downregulated in various hypertensive animal models (Kirsch et al., 2001; Grayson et al., 2007) as well as atherosclerotic plaques in nonhuman primates (Li et al., 2004) and humans (Adams et al., 2006). Furthermore, RGS5 seems to be an important regulator of vascular remodeling, under both normal and pathophysiological conditions (Armulik et al., 2005; Berger et al., 2005; Wang et al., 2016). For example, vascularization of tumors is normalized in RGS5 knockout mice, and immune-mediated destruction of solid tumors was more effective, resulting in greatly improved survival (Hamzah et al., 2008). This may be due to the role of RGS5 in pericytes (Cho et al., 2003), which critically support nascent vascularization (Mitchell et al., 2008). In this study, RGS5 inhibits multiple signaling pathways induced by angiotensin II, endothelin I, and others (Cho et al., 2003). RGS5-expressing pericytes have a demonstrated role in tumor angiogenesis (Bergers and Song, 2005; Ribeiro and Okamoto, 2015) and, together with previous work (Hamzah et al., 2008), strongly support a role for RGS5 in this disease model. Mouse models also show that blood pressure is reduced in RGS5-null mice (Cho et al., 2008; Nisanciglu et al., 2008). Accordingly, a human genome-wide screen found RGS5 expression was linked to blood pressure regulation (Chang et al., 2007). The role of RGS5 in vascular physiology and pathophysiology was recently reviewed (Ganss, 2015).

RGS8 appears to be enriched throughout the brain (Larminie et al., 2004), especially in Purkinje cells of the cerebellum (Gold et al., 1997; Saitoh and Odagiri, 2003; Saitoh et al., 2003). Interestingly, when RGS8 cDNA is expressed in non-neuronal cells, it accumulates in the nucleus, whereas in Purkinje neurons RGS8 is found within the soma and dendrites (Itoh et al., 2001). Coexpression with constitutively active Gao protein, or expression of RGS8 lacking the N terminus, reversed the nuclear localization (Saitoh et al., 2001, 2003), suggesting either robust nuclear export in Purkinje neurons of RGS8, or a cytosolic-localizing binding partner of RGS8 that is specific to Purkinje neurons versus non-neuronal cells. Furthermore, unlike canonical RGS function on G protein–gated potassium channels that accelerate channel desensitization (Chen et al., 2014; Ostrovskaya et al., 2014; Wydeven et al., 2014), RGS8 speeds up both the on-rate as well as the off-rate channel kinetics (Saitoh et al., 1997). RGS8 knockout mice have been generated, but no overt phenotype or histologic abnormality was found (Kuwata et al., 2008). Thus, RGS8’s role in physiology remains uncertain although likely important for key regulatory processes. Although not much is known regarding RGS8 links to human disease, one study found that electroconvulsive seizures in rats caused an increase in RGS8 mRNA in the prefrontal cortex 2 hours following acute shock and significantly reduced RGS8 mRNA in hippocampus 24 hours following both acute and chronic shock (Gold et al., 2002), suggesting a potential role for RGS8 in seizures.

RGS13 is another immune-specific modulator that is expressed in B and T lymphocytes (Shi et al., 2002; Estes et al., 2004), as well as mast cells (Bansal et al., 2008a,b). In B and T cells, RGS13 acts to desensitize chemokine receptor signaling (Shi et al., 2002; Estes et al., 2004; Han et al., 2006) similar to RGS1, and RGS13 knockout mice have enhanced B cell responses (Hwang et al., 2013). In mast cells, RGS13 constrains allergic responses generated by IgE (Bansal et al., 2008a,b), and RGS13 knockout mice exhibit enhanced IgE-mediated anaphylaxis. RGS13 transcript is greatly increased in adult T cell leukemia/lymphoma (Pismasion et al., 2009; Sethakorn and Dulin, 2013) and asthma (Raedler et al., 2015), underscoring both its role in immune cells and the importance of homeostatic balance of RGS13 signaling.

Originally cloned from the retina (Chen et al., 1996; Natochin et al., 1997), RGS16 has a relatively broad expression pattern, including the heart (Patten et al., 2002), brain (Grafstein-Dunn et al., 2001), liver (Kurrasch et al., 2004), and immune system (Beadling et al., 1999; Shi et al., 2004; Kveberg et al., 2005; Kim et al., 2006). RGS16 localization and GAP activity are regulated by addition of a palmitate at multiple cysteine residues (Hiol et al., 2003; Osterhout et al., 2003). As with many R4 family members, one of the most well-defined functions for RGS16 has been its role in adaptive immunity. RGS16 is involved in trafficking and migration of T lymphocytes (Estes et al., 2004) in response to an allergen challenge (Lippert et al., 2003), and RGS16 knockout mice have an enhanced inflammatory response in the lung (Shankar et al., 2012). Apart from its defined roles in allergic responses, RGS16 also has an intriguing role in regulating circadian systems (Goto et al., 2017). Within the brain, RGS16 is expressed in both the suprachiasmatic nucleus (SCN) and the thalamus (Grafstein-Dunn et al., 2001; Ueda et al., 2002). The SCN sets the global circadian clocks through cyclical signaling pathways, where RGS16 expression is cyclical. Loss of RGS16 causes dysregulation in circadian signaling in the SCN as well as delayed, shorter circadian behavioral activity in mice (Doi et al., 2011; Hayasaka et al., 2011). Another component of SCN circadian regulation is feeding behavior, which is synchronized with circadian rhythms coregulated by the liver. In mice, food anticipatory activity was found to be attenuated in RGS16 knockout mice during a restricted feeding schedule (Hayasaka et al., 2011). Complementary studies in the
liver showed that RGS16 knockout mice had higher rates of fatty acid oxidation, whereas mice that overexpress RGS16 had lower rates of fatty acid oxidation and higher blood triglyceride levels (Pashkov et al., 2011). Interestingly, in humans, noncoding variants in RGS16 have been linked with self-reported “morning people” (Hu et al., 2016; Jones et al., 2016). Finally, RGS16 expression has been linked with various cancers (Liang et al., 2009; Miyoshi et al., 2009; Carper et al., 2014).

RGS18 expression is mostly confined to bone marrow–derived cells (Nagata et al., 2001; Park et al., 2001; Yowe et al., 2001), more specifically in platelets (Gagnon et al., 2002). Although very little is known about RGS18 beyond its expression, mechanistically it seems to be important for regulating platelet activation (Gegenbauer et al., 2012; Ma et al., 2012). RGS18 knockout mice have reduced bleeding compared with wild-type mice, hyper-responsive platelet activation (Alshbool et al., 2015), and reduced platelet recovery following acute thrombocytopenia (Delesque-Touchard et al., 2014). In humans, RGS18 mRNA is elevated in aspirin-resistant platelets (Mao et al., 2014), and RGS18 protein was elevated in amyotrophic lateral sclerosis patients (Hagmark et al., 2014). Beyond its role in platelets, human SNVs near the RGS18 gene were associated with suicide attempts (Schosser et al., 2011), although no clear mechanism for this is known.

Among the R4 family, RGS21 is the smallest and most recently cloned RGS protein (von Buchholtz et al., 2004). It was originally cloned from bitter and sweet taste cells, although the protein may be expressed much more broadly (Li et al., 2005). In bitter taste cells, RGS21 was found to inhibit bitter taste signaling to cAMP, suggesting a role of RGS21 in the gustatory system (Cohen et al., 2012). Since then, human SNVs have linked RGS21 with celiac disease (Sharma et al., 2016). Nonetheless, roles for RGS21 in physiology and disease remain largely unexplored.

B. The R7 Family

1. R7 Family Overview. The R7 family of RGS proteins is composed of RGS6, RGS7, RGS9, and RGS11 (Fig. 1). These are highly homologous proteins mostly expressed in the nervous system, where they have a role in neuronal G protein signaling controlling nociception, reward behavior, motor control, and vision (Gold et al., 1997; Anderson et al., 2009; Gerber et al., 2016). The R7 RGS proteins contain distinctive domains that form stable stoichiometric heterotrimeric complexes with accessory binding partners that control protein–protein interaction, subcellular localization, and protein stability (Anderson et al., 2009; Sjögren, 2011). Besides the canonical RGS domain, other domains include the disheveled EGL10-Pleckstrin (DEP) homology domain, a R7 homology domain, and a G protein γ subunit-like (GGL) domain (Gold et al., 1997; Sjögren, 2011; Ahlers et al., 2016; Gerber et al., 2016). The RGS domain is located at the C terminus, where it stimulates GTP hydrolysis on Gai/o protein subunits (Snow et al., 1998b; Posner et al., 1999; He et al., 2000; Hooks et al., 2003; Martemyanov and Arshavsky, 2004; Anderson et al., 2009; Masuho et al., 2013; Stewart et al., 2015). The GGL domain, located upstream from the RGS domain, is structurally homologous to conventional γ subunits of G proteins (Posner et al., 1999; Anderson et al., 2009) and binds Gβ5 (type 5 G protein β subunit) as an obligatory partner (Anderson et al., 2009), which is crucial for protein stability (Snow et al., 1998b; Anderson et al., 2009; Sjögren, 2011; Gerber et al., 2016). Consistent with the brain expression patterns of R7 family members, various neurologic conditions such as anxiety, schizophrenia, drug dependence, and visual complications have been linked with the function of these proteins. Table 1 contains a brief overview of each protein within the R7 family, its tissue distribution, and its reported links to physiology and disease. Figure 1 shows a phylogenetic map of all RGS family proteins, with R7 family proteins highlighted in blue.

2. R7 Family Proteins in Human Physiology and Disease. RGS6 is highly expressed, at both the mRNA level and protein level, in brain tissue and in the heart (Gold et al., 1997; Ahlers et al., 2016). Within heart, RGS6 functions as an essential modulator of parasympathetic activation to prevent parasympathetic override and severe bradycardia (Yang et al., 2010). Studies relating RGS6 to human diseases are limited, although literature suggests that RGS6-specific modulation of Go may be involved in regulating several central nervous system diseases such as alcoholism (Stewart et al., 2015), anxiety and depression (Stewart et al., 2014), Parkinson’s disease (Bifsha et al., 2014), Alzheimer’s disease (Moon et al., 2015), schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics, 2014), and vision (Chograni et al., 2014). Prolonged exposure to alcohol upregulates RGS6 protein in a brain region known as the ventral tegmental area of wild-type mice. RGS6 knockout mice have reduced striatal dopamine, ameliorated alcohol-seeking behavior, and a reduction in alcohol-conditioned reward and withdrawal (Stewart et al., 2015). RGS6 knockout mice also showed protection from pathologic effects of chronic alcohol consumption on peripheral tissues, believed to be due to direct or indirect regulation by RGS6 of reactive oxygen species (Stewart et al., 2015; Ahlers et al., 2016). RGS6 is also enriched in dopaminergic neurons of the substantia nigra pars compacta, which are characteristically lost in Parkinson’s disease (Ahlers et al., 2016). Studies have shown that RGS6 knockout mice, but not wild-type mice, suffered from age-onset neurodegeneration of these neurons by the first year. These results also correlate with a decrease in
Gene products associated with differentiation and maintenance of dopamine neurons during development (Li et al., 2009; Ahlers et al., 2016), and whose expression is dysregulated in RGS6 knockout mice (Ahlers et al., 2016). Within mouse brain, RGS6 is also expressed in cortical and hippocampal neurons, where it mediates anxiety and depression (Stewart et al., 2015). RGS6 knockout mice displayed spontaneous anxiolytic and antidepressant behaviors that are sensitive to 5-HT1A receptor antagonism (Ahlers et al., 2016). Several studies have suggested RGS6 SNVs are significantly associated with multiple central diseases, including Alzheimer’s disease (rs4899412) (Moon et al., 2015; Ahlers et al., 2016) and schizophrenia (rs2332700) (Schizophrenia Working Group of the Psychiatric Genomics, 2014; Ahlers et al., 2016). Accordingly, the RGS6 gene can influence the pathophysiological processes underlying Alzheimer’s disease (Moon et al., 2015). Outside of the brain, atypical RGS6 protein expression is linked to several forms of cancer. A C→T SNV located in the 3’ untranslated region of the RGS6 gene was associated with a 34% reduction in bladder cancer and an increase in RGS6 protein (rs2074647) (Berman et al., 2004; Ahlers et al., 2016). RGS6 expression was found to be negatively correlated with human pancreatic cancer (Ahlers et al., 2016), human breast cancer progression (Maity et al., 2011, 2013; Ahlers et al., 2016), and resistance to chemotherapies (Maity et al., 2013). Finally, in roles unrelated to cancer, a splice mutation in RGS6 was identified as a genetic cause of autosomal recessive congenital cataract, mental retardation, and microcephaly in two Tunisian siblings (Chograni et al., 2014).

RGS7 is highly expressed in brain, in particular regions linked to anxiety such as the amygdala, hippocampus, brain stem, and hypothalamus (Larminie et al., 2004; Hohoff et al., 2009). RGS7 mRNA is found in abundance in neurons of ventral tegmental area and nucleus accumbens (Sutton et al., 2016), which have roles in drug reward and reinforcement. Within this circuit, the euphoric and analgesic effects of morphine are mediated by the μ-opioid receptor. RGS7 knockout mice showed an enhancement in reward behavior, increased analgesia, delayed tolerance, and heightened withdrawal in response to morphine administration (Sutton et al., 2016). Furthermore, chromosome 1q43, which contains the RGS7 gene, has been reported as a risk loci for panic disorder (Hohoff et al., 2009), and RGS7 SNV rs11805657 is associated with panic disorder (Crowe et al., 2001; Gelernter et al., 2001; Hohoff et al., 2009). Outside of the brain, a Genome-wide Association Study shows modest evidence for the involvement of RGS7 intron variants (rs4660010 and rs261809) in multiple sclerosis. Accordingly, it has been suggested that alteration in RGS7 function could potentially impair the normal dampening of the inflammatory response, leading to multiple sclerosis (McCauley et al., 2009). Overall, less is known about RGS7’s link to physiology and pathophysiology relative to other RGS, and the human variant information provided in this review may prove helpful in defining RGS7’s involvement in potential traits or disease.

The RGS9 gene forms two products: a short retina-specific transcript variant (RGS9-1) where it acts as a GAP on mammalian photoreceptor-linked G proteins (He et al., 1998), and a long brain-specific transcript variant (RGS9-2) enriched in the striatum (Zhang et al., 1999). The full-length human RGS9-1 consists of 484 amino acids, whereas RGS9-2 contains almost 200 extra amino acids at its C terminus (Zhang et al., 1999). Both RGS9 variants have a RGS domain and a DEP domain that binds to adaptor proteins such as R9AP (RGS9-1 anchor protein) in the retina, and R7BP (R7 binding protein) in the brain (Traynor et al., 2009). RGS9-1 accelerates the GTPase activity of transducin and colocalizes with other members of the phototransduction cascade (He et al., 1998; Zhang et al., 1999). During the recovery phase of visual transduction, RGS9 is anchored to a photoreceptor outer segment by R9AP, and mutations in both proteins have been associated with stationary retinal dysfunction syndrome, including RGS9 W299R (Nishiguchi et al., 2004; Cheng et al., 2007; Hartong et al., 2007; Stockman et al., 2008; Michaelides et al., 2010). RGS9-1 loss of function in the retina leads to bradyopsia, an inability to see moving objects during sudden changes in light intensity (Nishiguchi et al., 2004). Importantly, variants such as R128X (a nonsense mutation) in RGS9-1 create a truncated gene product lacking important domains crucial for a functional protein (Michaelides et al., 2010). RGS9-2 is expressed in the striatum of rat (Gold et al., 1997; Zhang et al., 1999) and human brain (Thomas et al., 1998; Zhang et al., 1999; Rahman et al., 2003; Liou et al., 2009), a region highly involved in antipsychotic-induced tardive dyskinesia. Genetic differences in RGS9-2 may play a role in patients developing tardive dyskinesia after antipsychotic treatment, because of RGS9-2 regulation of the D2 dopamine receptor (Rahman et al., 2003; Cabrera-Vera et al., 2004; Kovoor et al., 2005; Celver et al., 2010; Waugh et al., 2011). Intronic SNVs rs8077696, rs8070231, and rs2292593 were reported to likely alter binding efficiency of RGS9-2 to D2DR and play an important role in the development of tardive dyskinesia (Liou et al., 2009). RGS9-2 also regulates the μ-opioid receptor (Zachariou et al., 2003; Psifogeorgou et al., 2007, 2011; Waugh et al., 2011) and modulates reward responses through both opioid and dopamine receptors (Hoeks et al., 2008; Waugh et al., 2011). These neurotransmitter systems regulate feeding behavior and body weight (Bodnar, 2004; Gainetdinov, 2007; Waugh et al., 2011) in addition to reward response (Le Merrer et al., 2009; Johnson and Kenny, 2010; Waugh et al., 2011). A study showed an association between rs3215227
(an intronic variant) and significant higher body mass index in East Asian subjects (Waugh et al., 2011). RGS9-2 involvement in the dopamine reward pathway has suggested its involvement in addiction behavior. Cocaine self-administration in rats shows decreased RGS9-2 levels in the striatum compared with controls (Rahman et al., 2003; Traynor et al., 2009). RGS9 knockout mice have distinct locomotor-activating actions of dopaminergic or opioidergic agents such as cocaine, amphetamine, or morphine when compared with wild-type mice (Rahman et al., 2003; Blundell et al., 2008; Traynor et al., 2009). The absence of RGS9 also shows accelerated locomotor sensitization and increased reward sensitivity (Psifogeorgou et al., 2007; Blundell et al., 2008; Traynor et al., 2009).

RGS11 is highly expressed in the brain, especially in retinal bipolar and nerve cells (Rao et al., 2012; Cao et al., 2012; Shim et al., 2012), as well as outside the brain (Yang et al., 2016b). RGS11 interacts with mGluR6 at the dendritic tips of ON bipolar cells to regulate light-evoked responses (Cao et al., 2012). Gq5 knockout mice have greatly reduced RGS11 expression in the retina, resulting in dysfunctional photoreceptor signaling (Rao et al., 2007; Cao et al., 2012; Shim et al., 2012). Beyond the brain, recent studies have shown that upregulation of RGS11 might play a role in cancer. RGS11 is highly overexpressed in multiple tumors and associated with increased primary tumor status, nodal metastasis, and disease stage (Yang et al., 2016b). In colorectal cancer, RGS11 is upregulated and involved in chemotherapy resistance (Martinez-Cardus et al., 2009; Yang et al., 2016b). Although RGS11’s role in retinal bipolar and nerve cells has been described, its role in cancer and other diseases remains yet to be fully defined (Yang et al., 2016b).

C. The R12 Family

1. R12 Family Overview. The R12 family is a diverse group of RGS proteins, consisting of three members: RGS10, RGS12, and RGS14. Each has its own unique structure and function, but share a conserved RGS sequence and dynamic nuclear shuttling (Burgon et al., 2001; Chatterjee and Fisher, 2002; Cho et al., 2005; Waugh et al., 2005; Shu et al., 2007). Whereas RGS10 is a small, simple RGS protein that resembles the R4 family members, RGS12 and RGS14 have larger and more complex structures that share homology. Both RGS12 and RGS14 contain accessory domains, including two tandem Ras/Rap-binding domains (R1 and R2) and a G protein regulatory (GPR) motif. The R1 domains of RGS12 and RGS14 each interact with small G proteins such as Rap2 and H-Ras to regulate mitogen-activated protein kinase signaling (Traver et al., 2000; Willard et al., 2009; Shu et al., 2010; Vellano et al., 2013). The GPR motif binds inactive (as opposed to active GTP-bound) Ga proteins and serves as an inhibitor of GDP release (Kimple et al., 2001, 2002, 2004; Mittal and Linder, 2004) and also a regulator of RGS protein subcellular localization and membrane attachment (Shu et al., 2007; Brown et al., 2015b). RGS12 is expressed in humans as multiple splice variants (Chatterjee and Fisher, 2000), the longest of which (called trans-spliced, RGS12-TS) contains two additional domains: a PDZ domain and a PTB domain. PDZ domains are important regulators of localization and interaction with binding partners (Dunn and Ferguson, 2015). For example, RGS12-TS binds to CXCR2 via its PDZ domain (Snow et al., 1998a) as a means of directing to its target signaling partners. The PTB domain binds phosphorylated extracellular signal-regulated kinase (ERK) (Sambi et al., 2006). The demonstrated roles of these accessory domains are important to consider in RGS protein function/regulation beyond the canonical RGS domains highlighted in our review in this work. Table 1 contains a brief overview of each protein within the R12 family, its tissue distribution, and its reported links to physiology and disease. Figure 1 shows a phylogenetic map of all RGS family proteins, with R12 family proteins highlighted in gold.

2. R12 Family Proteins in Human Physiology and Disease. RGS10, at 20 kDa, is one of the smallest RGS family proteins, and is highly expressed in the brain and immune system (Gold et al., 1997; Haller et al., 2002). In humans, there are three splice variants of RGS10, differing by only a few amino acids at the N terminus. However, these small differences can have a substantial effect on RGS10 function, as the shortest splice variant (lacking only 14 amino acids) has impaired GAP activity (Ajit and Young, 2005). RGS10 is also dynamically regulated within the cell. Palmitoylation of an N-terminal cysteine targets RGS10 to the plasma membrane and enhances its GAP activity (Tu et al., 1999), whereas phosphorylation of a C-terminal serine targets RGS10 to the nucleus and impedes its GAP activity (Burgon et al., 2001). RGS10 has been documented in the nuclei of microglia and neurons (Waugh et al., 2005), where it may serve to regulate neuroinflammation. Indeed, RGS10 has been shown to promote survival of dopaminergic neurons via regulation of neuroinflammatory pathways in nigrostriatal circuits (Lee et al., 2008, 2011), implicating a neuroprotective role for RGS10 in dopaminergic disorders such as Parkinson’s disease (Tansey and Goldberg, 2010). Interestingly, a polymorphism (V38M or V44M in canonical sequence) in RGS10 was found in Japanese patients with schizophrenia, but it was not found to be significantly associated with disease due to sample size (Hishimoto et al., 2004). In peripheral immune cells, RGS10 regulates macrophage activation (Lee et al., 2013) and platelet activation (Hensch et al., 2016) and T lymphocytes (Lee et al., 2016), with potential roles in
clotting or autoimmune diseases. Additionally, loss of RGS10 in aged mice is linked with dysregulated peripheral immune cells and inflammatory cytokines (Kannarkat et al., 2015). Last, there is a curious link between RGS10 and chemoresistant ovarian cancer (Hooks et al., 2010; Ali et al., 2013; Caean et al., 2014; Hooks and Murph, 2015), potentially via a Rheb-GTP/mTOR pathway (Altman et al., 2015). A comprehensive review of the roles of RGS10 in neurons and immune cells was recently published (Lee and Tansey, 2015).

RGS12, in contrast to RGS10, is the largest RGS protein family member, with multiple splice variants ranging in size from 55 to 155 kDa. As outlined above, RGS12 contains additional signaling domains other than a RGS domain (Snow et al., 1997, 1998a; Ponting, 1999) that interact with various proteins. Beyond this, RGS12 also has been shown to interact with calcium channels in neurons (Schiff et al., 2000; Richman et al., 2005), and de novo mutations have been linked with schizophrenia (Xu et al., 2011). RGS12 expression has been reported throughout the body, including the brain, lung, testis, heart, and spleen (Snow et al., 1997; Doupnik et al., 2001). Like RGS10, RGS12 also shuttles in and out of the nucleus, where it has been shown to repress transcription (Chatterjee and Fisher, 2002; Lopez-Aranda et al., 2006). RGS12 is also expressed in osteoclasts and regulates differentiation (Yang and Li, 2007). Accordingly, RGS12 knockout mice have aberrant bone mass (Yang et al., 2013; Yuan et al., 2015b), suggesting a potential role in osteoporosis. Finally, RGS12 has been linked with cardiac hypertrophy (Huang et al., 2016a) and various cancers (Dai et al., 2011; Wang et al., 2017b).

RGS14 is a ∼60-kDa protein within the R12 family that is expressed in brain, heart, and spleen of rodents (Snow et al., 1997; Hollinger et al., 2001; Li et al., 2016). Although its brain expression pattern in adult rodents is largely limited to hippocampal area CA2, RGS14 has a wider brain distribution pattern in monkey and human brain (Squires et al., 2018), including multiple nuclei of the basal ganglia. Of note, within striatum of monkey brain, RGS14 appears to express several shorter splice variants not observed in rodents (Squires et al., 2018). Although RGS14 has not been conclusively linked with any specific diseases in the brain, a Genome-wide Association Study identified RGS14 as a risk factor for multiple sclerosis (Ryu et al., 2014), and a follow-up mouse study confirmed differential expression in a mouse model of multiple sclerosis (Sevastou et al., 2016). RGS14 also suppresses hippocampal-based synaptic plasticity and learning in the CA2 region of the hippocampus in mice (Lee et al., 2010), which has been linked to social and contextual memory (Hitti and Siegelbaum, 2014; Alexander et al., 2016; Dudek et al., 2016; Piskorowski et al., 2016). Although RGS14’s role in these behaviors has not yet been fully elucidated (Evans et al., 2015), its high level of expression in area CA2 suggests it may be a key regulator of hippocampal-based learning and memory. Outside of the hippocampus, RGS14’s role within the basal ganglia suggests a link to movement disorders, such as Parkinson’s disease, and transcriptional studies of Parkinson’s patients showing decreases in RGS14 mRNA support a possible role in this process (Vogt et al., 2006). In the periphery, RGS14 expression is downregulated in failing human hearts, which suppresses cardiac remodeling through regulation of the mitogen-activated protein kinase kinase/ERK pathway (Li et al., 2016). RGS14 interacts with active H-Ras-GTP and Raf-1 and to block ERK signaling (Willard et al., 2009; Shu et al., 2010; Vellano et al., 2013), and RGS14 actions on cardiac remodeling presumably are mediated through one or both Ras/Rap-binding domains on RGS14 (Li et al., 2016), highlighting the importance of accessory domains on RGS protein functions independent of the canonical RGS domain. Finally, multiple genetic studies have found variants in the proximity of the RGS14 gene that are associated with kidney disease (Urabe et al., 2012; Yasui et al., 2013; Mahajan et al., 2016) and altered serum concentrations of both parathyroid hormone (Robinson-Cohen et al., 2017) and phosphorous (Kestenbaum et al., 2010), implicating a potential role for RGS14 in regulating the homeostasis of serum phosphate and other ions.

D. The RZ Family

1. RZ Family Overview. The RZ family is composed of RGS17, RGS19, and RGS20. These all are small, simple RGS proteins similar to the R4 family members. However, unique to the RZ family members is a conserved string of cysteine residues found near their N termini that is palmitoylated and regulates both their membrane localization and interaction with binding partners (De Vries et al., 1996; Nunn et al., 2006). RZ proteins also function as adapter proteins for Gα subunit degradation and play important roles in the regulation of signaling and cytoskeletal events in the brain (Mao et al., 2004). They are also highly conserved in metazoans and most closely related to the R4 RGS family (Sierra et al., 2002; Nunn et al., 2006). All members of this family can bind to certain members of the Gαi and Gαq subfamily, but with some selectivity (Tu et al., 1997; Glick et al., 1998; Mao et al., 2004). Table 1 contains a brief overview of each protein within the RZ family, its tissue distribution, and its reported links to physiology and disease. Figure 1 shows a phylogenetic map of all RGS family proteins, with RZ family proteins highlighted in red.

2. RZ Family Proteins in Human Physiology and Disease. RGS17, also known as RGSZ2, demonstrates GAP activity for Gαi/o, Gαz, and Gαq (Mao et al., 2004). In humans, RGS17 is expressed in the nucleus accumens, hippocampus, and putamen, with highest
expression found in the cerebellum (Mao et al., 2004; Hayes and Roman, 2016). However, outside the brain, RGS17 has been reported to be overexpressed in human lung adenocarcinomas and prostate cancer (Mao et al., 2004; James et al., 2009; You et al., 2009). In lung, colon, and prostate tumor cell lines, knocking down RGS17 results in decreased tumor growth and tumor cell proliferation; conversely, overexpression of RGS17 in these cell lines resulted in increased tumor growth (James et al., 2009; You et al., 2009). Underscoring this, SNVs in the first intron of the RGS17 gene (rs6901126, rs4083914, and rs9479510) are associated with lung cancer (You et al., 2009). RGS17 also is overexpressed in human liver cancer (Hayes and Roman, 2016). Furthermore, when ovarian cancer cells are treated with chemotherapeutic agents, there is a loss of RGS17 expression, suggesting a role for RGS17 in chemoresistance, perhaps by promoting cell survival via phosphatidylinositol 3-kinase/AKT signaling in these cells (Hooks et al., 2010; Hayes and Roman, 2016). Outside of RGS17 links to cancer, postmortem brain samples from patients with clinical depression show a decrease in RGS17 expression (Shelton et al., 2011; Hayes and Roman, 2016). Furthermore, SNVs in RGS17 found in the promoter region (rs596359) and introns (rs6931160, rs9397585, rs1933258, rs9371276, rs516557, and rs545323) are associated with substance dependence (Zhang et al., 2012; Hayes and Roman, 2016). Finally, two intronic RGS17 SNVs are associated with smoking initiation, rs7747583 and rs2349433 (Yoon et al., 2012; Hayes and Roman, 2016). Whereas possible mechanisms are unknown, expression of RGS17 in brain regions known to be linked to substance dependence supports the relationship between SNVs and these diseases.

Comparatively less is known about RGS19, which is highly expressed (by mRNA) in the heart, lung, and liver, but very low in brain, and seems to regulate proliferation of embryonic stem cells (De Vries et al., 1995; Ji et al., 2015). Unlike the other two RZ family members, RGS19 preferentially interacts with Gα3 (De Vries et al., 1995). One report showed loss of RGS19 slightly enhances opioid-induced analgesia at µ-opioid receptors (Garzon et al., 2004), a brain-specific function that suggests even low expression can have a functional impact. RGS19 expression levels are reportedly upregulated in several disease states, including multiple sclerosis (Igci et al., 2016) and ovarian cancers (Tso et al., 2011). RGS19 is also highly expressed in human neuroblastoma SH-SY5Y cells (Wang and Traynor, 2013). However, in other instances, RGS19 has been reported to inhibit Ras activation by upregulating Nm23, a tumor metastasis suppressor (Wang et al., 2013). Transgenic mice overexpressing RGS19 exhibited multiple heart defects during development and increased expression of heart failure–related biomarkers, including B-type natriuretic peptide and β-major histocompatibility complex (Ji et al., 2010). Although there are few reports to date defining RGS19 in human disease, the overexpression studies in mice highlight the potential heart disease contribution of GoF variants within the coding region.

RGS20, also known as RGSZ1 and Ret RGS, selectively interacts with Gαz and Gα2 subunits (Wang et al., 1998, 2002). The RGS20 transcript is highly expressed in human caudate nucleus and temporal lobe (Wang et al., 1998), and RGS20 splice variants are detectable in the eye (Barker et al., 2001; Yang et al., 2016a). Loss of RGS20 in mice leads to enhanced µ-opioid–induced analgesia and tolerance to morphine (Garzon et al., 2004). Outside the brain, RGS20 is found at significantly high levels in melanoma and metastatic breast cancer cells (Yang et al., 2016a). Furthermore, expression of RGS20 in HeLa, breast adenocarcinoma MDA-MB-231, non-small cell lung carcinoma H1299, and A549 cells results in enhanced cell aggregation, migration, invasion, and adhesion, suggesting a role for RGS20 in tumor metastasis (Yang et al., 2016a). A recent study on triple-negative breast cancer reported RGS20 was overexpressed in those tissues, and that protein expression correlated with disease progression/prognosis, suggesting a novel target for therapy (Li et al., 2017). Finally, RGS20 is reported to be significantly associated with hypertension, where it may synergistically interact with other genes to predispose patients to hypertension (Kohara et al., 2008).

### III. Analysis of Rare Human Variants of RGS Proteins

As outlined above, RGS proteins play key roles in human physiology and disease, and rare human variants are thought to underlie many complex human diseases and traits. Therefore, we took advantage of recently available human exome sequencing databases and newly described bioinformatics/proteomic analytical tools to identify rare human variants in canonical RGS proteins that we predict will have a marked CoF phenotype. Using knowledge (published structural Protein Data Bank files) of the binding interface between the RGS domain and partner Gα subunit, we focus on individual variants of interest derived from these vast datasets that are likely to disrupt RGS-Gα interactions. Whereas a LoF or GoF may lead to disease states, a GoF in one pathway may in fact lead to a LoF in a completely separate pathway. Thus, as described above, we consider a CoF that accounts for either LoF or GoF. We postulate that these variants may be missed by genome association studies due to their rarity (1%–2% or less), but nonetheless confer the same phenotype (i.e., multiple genotypes may have the same phenotype), or may redirect the RGS function (e.g., by mislocalization) to affect atypical pathways. Given these caveats, in this
Residues were selected based on our criteria (outlined in the text) of CADD >20, residues that overlap with conserved sites important for Gα interaction (determined by structural insights), and PTM alignment analysis. Each residue mutation (native, residue position, mutation) has a description as to why it is predicted to disrupt function (“LoF Rationale”), its associated CADD depletion score (“CADD”), associated genic intolerance score (“MTR”), number of reported PTMs, number of neighboring PTMs (where a PTM is found within seven amino acids before/after), and prevalence in the population (“Prevalence”). Note that if the human variants are reported in the noncanonical transcript, the analogous residue position for the canonical sequence is coded in gray.

<table>
<thead>
<tr>
<th>RGS Protein</th>
<th>Residue Mutation</th>
<th>LoF Rationale</th>
<th>CADD</th>
<th>MTR</th>
<th>PTM</th>
<th>Neighboring PTMs</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RGS4</strong></td>
<td>L170(73)P</td>
<td>Proline in α helix</td>
<td>23</td>
<td>1.01</td>
<td>0</td>
<td>0</td>
<td>0.0043% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>E214(117)K</td>
<td>Salt bridge partner</td>
<td>27.3</td>
<td>0.50</td>
<td>0</td>
<td>0</td>
<td>0.0063% (Latino)</td>
</tr>
<tr>
<td></td>
<td>D227(130)G</td>
<td>100% conserved, salt bridge partner</td>
<td>29.6</td>
<td>0.90</td>
<td>0</td>
<td>0</td>
<td>0.0065% (African)</td>
</tr>
<tr>
<td></td>
<td>R231(134)W</td>
<td>Highly conserved, stabilizes switch III</td>
<td>29.8</td>
<td>1.04</td>
<td>0</td>
<td>0</td>
<td>0.0065% (African)</td>
</tr>
<tr>
<td></td>
<td>D260(163)N</td>
<td>Highly conserved, salt bridge partner, very high CADD</td>
<td>32</td>
<td>1.19</td>
<td>0</td>
<td>0</td>
<td>0.0065% (African)</td>
</tr>
<tr>
<td></td>
<td>D260(163)G</td>
<td>Highly conserved, salt bridge partner</td>
<td>29.4</td>
<td>1.19</td>
<td>0</td>
<td>0</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>R263(166)C</td>
<td>Contact with α helical domain of Go11</td>
<td>28.4</td>
<td>1.07</td>
<td>0</td>
<td>0</td>
<td>0.0116% (East Asian)</td>
</tr>
<tr>
<td></td>
<td>R263(166)H</td>
<td>Contact with α helical domain of Go11</td>
<td>24.5</td>
<td>1.07</td>
<td>0</td>
<td>0</td>
<td>0.0032% (Global)</td>
</tr>
<tr>
<td></td>
<td>R264(167)C</td>
<td>Highly conserved, salt bridge partner, very high CADD</td>
<td>35</td>
<td>1.06</td>
<td>0</td>
<td>0</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td><strong>RGS9</strong></td>
<td>W299R</td>
<td>Link to disease, participates in electrostatic interaction</td>
<td>29.1</td>
<td>0.71</td>
<td>0</td>
<td>1</td>
<td>0.0229% (European)</td>
</tr>
<tr>
<td></td>
<td>R364C</td>
<td>Stabilizes α5-α6 loop of RGS domain</td>
<td>29.5</td>
<td>0.78</td>
<td>0</td>
<td>0</td>
<td>0.0018% (European)</td>
</tr>
<tr>
<td></td>
<td>K400Q</td>
<td>100% conserved, contact with α helical domain</td>
<td>28.1</td>
<td>0.97</td>
<td>0</td>
<td>0</td>
<td>0.0018% (European)</td>
</tr>
<tr>
<td></td>
<td>R406C</td>
<td>100% conserved, salt bridge partner, very high CADD</td>
<td>35</td>
<td>0.98</td>
<td>0</td>
<td>1</td>
<td>0.0030% (Latino)</td>
</tr>
<tr>
<td></td>
<td>R406H</td>
<td>100% conserved, salt bridge partner, very high CADD</td>
<td>34</td>
<td>0.98</td>
<td>0</td>
<td>1</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td><strong>RGS10</strong></td>
<td>L48(38)P</td>
<td>Proline in α helix</td>
<td>29.6</td>
<td>0.78</td>
<td>0</td>
<td>2</td>
<td>0.0032% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>V52(44)M</td>
<td>Possible link to schizophrenia, hydrophobic core partner</td>
<td>34</td>
<td>0.78</td>
<td>0</td>
<td>1</td>
<td>1.9239% (East Asian)</td>
</tr>
<tr>
<td></td>
<td>D141(133)N</td>
<td>Highly conserved, ionic stabilization, very high CADD</td>
<td>35</td>
<td>0.90</td>
<td>0</td>
<td>2</td>
<td>0.0183% (Other)</td>
</tr>
<tr>
<td></td>
<td>R145(137)C</td>
<td>Highly conserved, ionic stabilization, very high CADD</td>
<td>35</td>
<td>0.86</td>
<td>0</td>
<td>2</td>
<td>0.0617% (East Asian)</td>
</tr>
<tr>
<td></td>
<td>K148(140)R</td>
<td>Ubiquitination</td>
<td>25</td>
<td>0.81</td>
<td>1</td>
<td>1</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td><strong>RGS17</strong></td>
<td>E148G</td>
<td>100% conserved, likely ionic interaction</td>
<td>28.2</td>
<td>0.87</td>
<td>0</td>
<td>0</td>
<td>0.0010% (European)</td>
</tr>
<tr>
<td></td>
<td>P166L</td>
<td>100% conserved within and across family, high CADD</td>
<td>28.3</td>
<td>0.73</td>
<td>0</td>
<td>1</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>R189T</td>
<td>100% conserved, likely ionic interaction</td>
<td>28.1</td>
<td>1.03</td>
<td>0</td>
<td>0</td>
<td>0.0009% (European)</td>
</tr>
</tbody>
</table>

In this study we analyze the entire coding sequence for all 20 canonical RGS proteins (Supplemental Material) and provide a detailed case study of a representative RGS protein from each subfamily that we highlight in the main body of this review (RGS4 from R4, RGS9 from R7, RGS10 from R12, and RGS17 from RZ) (Figs. 2–6).

Selection of rare variants that have potential to produce CoF phenotypes was based on several factors described in this work. First and foremost, we identified rare variants as those for which prevalence is well below 2% in a given population (for this dataset: minimum = 0.0009%, maximum = 1.92%, median = 0.0032%). Second, we used combined annotation-dependent depletion (CADD) analysis, presented as a C score (Supplemental Data; Table 2), to estimate the potential deleteriousness of each variant—a feature that strongly correlates with both molecular functionality and pathogenicity (Kircher et al., 2014). A CADD C score above 20 (top 1%), which we used as a hard filter, indicates a very high likelihood of deleteriousness (for this dataset: minimum = 23, maximum = 35, median = 29.4). Third, we considered structural data and sequence conservation when available (Figs. 2A, 4A, 5A, and 6A). Variants that overlap with identified contact points between the RGS domain and Go provide a strong case for CoF and were selected. Similarly, variants that overlap with highly conserved residues (particularly those that reside in critical regions of the RGS domain as determined by high-resolution structural data) were selected. This criterion was particularly important for eliminating variants that fall on highly variable residues within the family (i.e., if the amino acid properties are not conserved within the family, it is likely a noncritical residue). Fourth, we used a newly described bioinformatics tool to measure the tolerance of each position in each RGS protein for mutation, expressed as a MTR (Figs. 2B, 4B, 5B, and 6B; Supplemental Material; Table 2), which estimates the functional sensitivity of a given residue to mutation (Traynelis et al., 2017). MTR represents a novel measure of purifying selection acting on missense variants in a 31-codon sliding window across the sequence of a gene. Neutrality, or a MTR of 1.0, represents the point at which the observed number of missense variants is equivalent to the expected number in the sliding window. MTR is a useful tool for interpreting missense protein variants in the
context of monogenic diseases (Traynelis et al., 2017) and, although may be less predictive for proteins linked to polygenic diseases, is nonetheless reported in this work (for this dataset: minimum = 0.5, maximum = 1.19, median = 0.94) (Supplemental Material; Table 2).

As a further level of analysis, we consider sites of PTM found in human RGS proteins that, in many cases, overlap with rare variants that exhibit high C scores (Figs. 2B, 4B, 5B, and 6B). As outlined above, PTMs serve critical regulatory roles in protein function and are often overlooked or missed in disease assessment (Torres et al., 2016). A recent advance in the analysis of PTMs (Dewhurst et al., 2015; Torres et al., 2016; Dewhurst and Torres, 2017) offers a powerful bioinformatics/proteomics tool for characterizing experimentally verified PTMs in the context of their alignment within protein or domain families [i.e., modified alignment positions (MAPs)] (see Supplemental Information and Hunter et al. [2009], Edgar [2004], and Hornbeck et al. [2015]) and provides a functionally impactful analysis that is complementary to CADD and MTR data. This approach has shown further promise in the discovery of PTMs associated with disease-linked mutations (Torres et al., 2016). Therefore, we surveyed the coincidence of PTMs/MAPs and mutations within each RGS protein, with the goal of highlighting positions where CoF might be attributable to a change in PTM status. For this purpose, PTMs unique to each human RGS protein are superimposed onto the MTR plots where the specific type of modification (Ph = phosphorylation, Ub = ubiquitination, Ac = acetylation, Pm = palmitoylation) is noted above the graph, and the median, 25%, and 5% MTR values are indicated by horizontal red, gray, and green lines, respectively, for representative family members (Figs. 2B, 4B, 5B, and 6B; Supplemental Material; Supplemental Table 1). In each case, only positions found to be modified in the given human RGS protein are shown (Figs. 2–6, red circles). Within the RGS domain of each protein, the size of each PTM site reflects the total number of PTMs observed within the domain MAP (see Supplemental Material), whereas positions outside the RGS domain simply indicate the position of a single PTM for the given protein. Each PTM site is further annotated to indicate when there is evidence of function for the PTM in the given protein (green outer ring) or evidence for function within the MAP for the domain family (red outer ring). Disease-linked variant positions are also indicated (orange circles). Although not included in our primary analysis, we also report neighboring PTM count (Table 2), which is a summation of reported PTMs found within a ±7 residue window surrounding the human variant. Variants that fall within this window may in fact conflict with the ability of a modifying enzyme to dock on the target protein.

Fig. 2. The R4 subfamily of RGS proteins: analysis of RGS4 human variants. (A) Clustal W was used to align sequences of the RGS domains within the R4 subfamily of human RGS proteins. The RGS4:Gαi1-AIF2 crystal structure (Protein Data Bank: 1AGR) was used to extract residues on RGS4 critical for interaction with Gαi1 (highlighted in yellow). We then added pale yellow-filled boxes to extend this information to all R4 subfamily members. (B) MTR and PTM cluster analysis plot for RGS4-3 (UniProt ID P49798-3; see text for details). The RGS domain of RGS4 (highlighted in blue) is under selective pressure, indicated by a low MTR (median, red line; 25%, gray dashed line; 5%, green dashed line). RGS4 is palmitoylated on three residues, and although one is found to be mutated in the human population (GnomAD version 2.0), it is not located within the RGS domain and therefore not highlighted in this analysis (but see Supplemental Material; Supplemental Table 1) (C) Rare human variants of RGS4 plotted along the sequence using Lollipops (https://github.com/pbnjay/lollipops). Missense (amino acid change) variants are displayed on top in teal, whereas silent (no amino acid change) variants are displayed on the bottom in purple. (D) From our analyses, we selected nine variants in the RGS domain predicted to have a CoF phenotype and mapped them onto the RGS4 sequence using Lollipops (https://github.com/pbnjay/lollipops). (E) These selected variants are mapped onto the reported crystal structure (Protein Data Bank: 1AGR) of RGS4 (gray, selected residues in red) in complex with Gαi1-AIF2 (light blue).
therefore indicate in our analysis below when this occurs and propose this may be one mechanism for a CoF phenotype. A comprehensive table of each experimentally observed PTM for all human RGS proteins is also included in Supplemental Table 1. Notably, in each RGS protein case study we present, the RGS domain is both under mutational selective pressure and a hotbed for PTM activity, supporting the idea that disruption of these domains by mutation is likely to alter critical cellular functions.

Taking all of this information into account, we plot the genetic distribution of both missense (top of Figs. 2C, 4C, 5C, and 6C) and silent (bottom of Figs. 2C, 4C, 5C, and 6C) variants for each RGS protein, along with a mapping of the selected rare variants of interest onto the protein sequence (Figs. 2D, 4D, 5D, and 6D). (Note: Human RGS protein exome sequences selected for analysis and presentation in this review are identified by UniProt IDs (Bairoch et al., 2005) listed in Supplemental Material and in the figure legends.) These selected variants, identified from the predictive criteria, are finally placed onto the reported crystal structure of the RGS protein (Figs. 2E, 4E, 5E, and 6E) and listed in Table 2. Importantly, whereas our focus and discussion are centered on the RGS domain of each protein subfamily, we present the comprehensive set of missense variants, MTR, CADD C score, and PTMs for the entire sequence of all 20 canonical RGS proteins (see Supplemental Material). Lastly, we validate this overall approach by testing one of the selected variants for RGS4 to show that this variant, as predicted, results in a profound CoF phenotype (Fig. 3). We suggest that this overall approach and dataset can be used and expanded to other individual RGS proteins across each family (and any protein family in general) to prioritize rare variants and PTMs for CoF studies and improved understanding of human pathophysiology.

A. R4 Family: RGS4 Rare Variants

Among the R4 subfamily of RGS proteins, RGS4 is perhaps the best characterized with abundant functional information, including the first solved RGS protein crystal structure (Tesmer et al., 1997). Therefore, RGS4 is well suited as a case study for analyzing and predicting CoF human variants within the R4 family. Consistent with this idea, a human mutation in RGS4, S30C, has previously been reported to display a GoF phenotype (enhanced GAP activity) that translated to RGS16 when an analogous mutation was made (Hill et al., 2008). This demonstrates the idea that insights gained from mutational analysis in one RGS protein may extend to other close family members, although in this study we emphasize variants unique to each individual RGS protein (Supplemental Material).

Human RGS4 has five reported splice variants with predicted sizes ranging from 93 to 304 amino acids (Ding et al., 2007). The RGS4 variants in the GnomAD database (Lek et al., 2016) are reported with respect to the longest splice variant (RGS4-3; UniProt ID: P49798-3), which adds almost 100 amino acids to the N terminus of the commonly used canonical human RGS4 reference sequence (RGS4-1; 205 amino acids; UniProt ID: P49798-1) (Ding et al., 2007). Therefore, for our purposes in this work, we list first the reported variant amino acid location within the longer RGS4-3 sequence, followed by the canonical residue position in parentheses (Table 2). In our analysis of RGS4, we first compare an alignment of all R4 family members and identify residues that are contact points for Ga3 highlighted in yellow, as determined by X-ray crystallography (Tesmer et al., 1997) (Fig. 2A). Conserved RGS:Ga3 interface contact regions across R4 family members are highlighted by boxes. Next, we examine MTR for our case study protein, RGS4 (Fig. 2B), with reported PTMs unique to human RGS4 plotted onto the MTR data and identified above. From this analysis, it is clear that the RGS domain of RGS4 is under selective pressure, suggesting that human variants in this region are likely to change the function of the protein. Thus, we take advantage of the reported crystal structure for RGS4 in complex with Ga31 (Tesmer et al., 1997), extracting information about contact points between the two proteins. From these structural, PTM, and CADD analyses, we identified nine variants, including the following: E214(117)K, D227(130)G, D260(163)N, D260(163)G, R264(167)C, L170(73)P, R231(134)W, R263(166)C, and R263(166)H (Fig. 2, C and D; Table 2). Of these, E214(117)K (C score 27.3), D227(130)G (C score 29.6), D260(163)N (C score 32), D260(163)G (C score 29.4), and R264(167)C (C score 35) are all highly or completely conserved residues among the R4 family (Fig. 2A) and participate in salt bridge interactions. Mutation of these residues to a noncharged [as with variants D227(130)G, D260(163)N, and D260(163)G], or opposite charge [as with E214(117)K] residue is predicted to decrease the stability of a RGS4:Ga3 complex. L170(73)P (C score 23) is located within an α helix and is, therefore, predicted to disrupt secondary structure of the RGS domain. Alanine acid R231(134) is another highly conserved residue within the family that participates in stabilizing switch III on the Ga3 subunit, an important interaction for promoting GTP hydrolysis. As such, the R231(134)W (C score 29.8) variant very likely disrupts RGS4 capacity to interact with Ga3 (see below). Finally, R263(166)C (C score 28.4) and R263(166)H (C score 24.5) are predicted to disrupt contact between the RGS domain and the all helical domain of Ga3. In these cases, disruption of binding to Ga3, or disruption of RGS4 capacity to stabilize switch regions on Ga3, should similarly affect downstream G protein signaling events and disease states. Table 2 lists human variants in RGS4 that are predicted to disrupt RGS4 interaction with Ga3, and/or GAP activity.

Further analysis of the RGS4 sequence using PTM alignment analysis identified three residues in human
RGS4 that undergo PTM. C99 and C109, both located outside the RGS domain, are palmitoylated. Of note, C109 corresponds to reported human variant C109(12)R and is an important regulator of RGS4 subcellular localization and GAP activity (Bastin et al., 2012). Therefore, this human variant, although not within our focus in this study on the RGS domain, is predicted to disrupt the function of RGS4 and thus may mimic loss of RGS4, as is the case for multiple diseases described above. In addition, C192(95), located within the RGS domain of RGS4, is palmitoylated and serves an important role in regulating RGS4 membrane localization and its GAP activity (Tu et al., 1999). However, this cysteine residue does not correspond with a reported human variant from this dataset. Thus, we mapped our selected rare variants for RGS4 onto the sequence (Fig. 2D) and structure (Fig. 2E) of RGS4.

In all, we identified nine variants (corresponding to seven residues) within the RGS domain of RGS4 that are predicted to exhibit CoF phenotypes. To test this prediction and validate our approach, we selected one human variant in RGS4 from our list, R231(134)W, to test for a CoF phenotype. Compared with wild-type RGS4, R231(134)W (Fig. 3A) exhibited a loss of direct Go1-AlF4 binding (Fig. 3B), as well as a dramatic decrease in Go1-YFP binding in live cells by bioluminescence resonance energy transfer analysis (Brown et al., 2015a), indicating a profound LoF phenotype (Fig. 3C). These data suggest that the capacity of the RGS domain of RGS4 to bind Go, and thus serve as a GAP, is disrupted by this naturally-occurring human variant, and may therefore be a determinant in RGS4-linked diseases found in carriers. For example, multiple neurologic diseases, including schizophrenia and alcoholism, correlate with low RGS4 expression (Mirnics et al., 2001; Ho et al., 2010), a phenotype that would also be expected for a LoF variant in the RGS domain of RGS4. Therefore, one possible hypothesis is that LoF variant R231(134)W is a risk factor for schizophrenia—a classification that could be difficult to make through unbiased genome-wide studies due to the extremely low prevalence of the variant in the population (0.0065% of the African population). Indeed, we propose that hypotheses generated through an integrated bioinformatics analysis such as this one could reveal several such cases wherein disease states can be more confidently linked to discrete changes in protein structure and function. Other LoF variants listed in Table 2 are predicted to share the same phenotype and outcome. Below, we take a similar approach to extend our analyses to other RGS subfamilies.

B. R7 Family: RGS9 Rare Variants

Unlike the R4 family, the R7 family of RGS proteins is all larger multidomain proteins (Fig. 1). Among the R7 subfamily of RGS proteins, RGS9 was chosen as a representative family member to analyze because of its solved protein crystal structure in complex with Go and known RGS9 links to disease (Slep et al., 2001; Nishiguchi et al., 2004; Cheng et al., 2007; Hartong et al., 2007; Stockman et al., 2008; Michaelides et al., 2010). Structural insight for the RGS9 RGS domain cocrystallized in complex with a Go1/t chimera reveals a mostly charged interface (Slep et al., 2001), similar to other RGS:Go interfaces, and provides information about residues that are both highly conserved and crucial for interaction (highlighted in Fig. 4A). The RGS9 MTR plot (Fig. 4B) demonstrates that the RGS domain, particularly the C-terminal half of the domain, is under selective pressure. This information indicates that the latter half of the domain is predicted to be sensitive to missense variation.

We next examined whether reported human PTMs aligned with human variants (Fig. 4B). Although there are many PTMs outside of the RGS domain, two rare-variant phosphorylation sites were found within the
RGS domain, S304R and Y413C. Whereas Y413C has a CADD C score of 27.2, S304R has a CADD C score of 3.7, which did not meet our cutoff. We mapped the rare human variants to the sequence of RGS9 (Fig. 4C) and noticed a large degree of variation across the gene, suggesting a great deal of sequence diversity. Based on the PTM, CADD, and structural analyses, we selected six variants of interest (W299R, R364C, K400Q, R406C, R406H, and Y413C; Table 2) and mapped them onto the RGS9 sequence (Fig. 4D) and structure (Fig. 4E). Most variants map to the C-terminal half of the RGS domain (Fig. 4B), with the exception of W299R, which was found in the literature to underlie some cases of retinal dysfunction (Nishiguchi et al., 2004; Cheng et al., 2007; Hartong et al., 2007; Stockman et al., 2008; Michaelides et al., 2010). Based on the crystal structure of RGS9 (Slep et al., 2001), W299 participates in an electrostatic interaction with a nearby lysine (K311) to stabilize the tertiary structure of the RGS domain, and mutation to a positively charged arginine in the W299R variant (C score 29.1) could disrupt this tertiary structure. As noted above, we report human variants that fall within a ±7 residue window of PTM. As such, W299R has one neighboring PTM, a phosphorylation on S304. Thus, one possibility is that the W299R variant leads to retinal dysfunction by disrupting phosphorylation of S304, a prospect that must be further explored. Y413C, although modified itself, also falls within a neighboring PTM window, an acetylation on K419. R364C (C score 29.5) makes a water-mediated contact with Gaαi1/t and stabilizes the α5-α6 loop, which undergoes the greatest conformational change when serving as a GAP for Ga. K400Q (C score 28.1) is perfectly conserved within the family, form a cis salt bridge with a nearby aspartate (D402) on RGS9 to help stabilize Ga switch I, and these variants have very high CADD C scores (35 and 34, respectively). R406C/H also falls within a neighboring PTM, that of Y413C.

Due to the known link between W299R and retinal dysfunction, we propose that these as-yet undefined variants listed in this study may have a similar role in retinal or other RGS9-linked diseases. Elsewhere within the CNS, RGS9-2 has a role in the striatum (Rahman et al., 1999) regulating dopamine signaling, which has consequences for addiction, Parkinson’s disease, and schizophrenia (Rahman et al., 2003; Kovoor et al., 2005; Seeman et al., 2007). Notably, these phenotypes, such as enhanced sensitivity to stimulants, arise from a loss of RGS9-2 protein. Thus, LoF variants described in this dataset may contribute to susceptibility of these diseases in carriers. Furthermore, as the R7...
subfamily proteins contain additional signaling domains that are critical for protein stability and proper subcellular localization (Anderson et al., 2007, 2009), LoF variants in those domains could also cause LoF phenotypes that should be explored further. Table 2 lists human variants in the RGS domain of RGS9 that we predict disrupt RGS9 interaction with Ga and/or GAP activity.

C. R12 Family: RGS10 Rare Variants

Among the R12 subfamily of RGS proteins, RGS10 is the smallest and simplest, and is the only family member that has been crystallized in complex with a Ga, Ga\textit{i3} (Soundararajan et al., 2008), providing an excellent case study for the R12 family. We first compared amino sequence conservation and highlighted contact points between RGS10 and Ga\textit{i3} (Fig. 5A). As with RGS4, human RGS10 has three splice variants that range in size from 167 to 181 amino acids (Rivero et al., 2010; Ali et al., 2013; Lee and Tansey, 2015). The RGS10 variants in the GnomAD database (Lek et al., 2016) are reported with respect to the longest splice variant (RGS10-3; UniProt ID O43665-3) and mapped onto this sequence in Fig. 5C, which adds only eight amino acids to the N terminus of the commonly used canonical human RGS10 reference sequence (RGS10-1; Uniprot ID O43665-1; 173 amino acids).

Compared with other solved RGS domain-G\textalpha crystal structures (Tesmer et al., 1997; Slep et al., 2001), the RGS domain of RGS10 makes fewer contact points with Ga (compare with highlighted regions in Fig. 4A; Fig. 2A), yet five variants were found in our predictive CADD/PTM/structural analysis: L46(38)P, V52(44)M, D141(133)N, R145(137)C, and K148(140)R (mapped onto the RGS10 sequence in Fig. 5D and RGS10 structure in Fig. 5E). The RGS10 MTR plot shows a high degree of selective pressure across the entire RGS domain (Fig. 5B). Furthermore, there are several reported human PTMs within the RGS domain, including K53 and K148 (ubiquitination), K78 (acyetylation), and Y94 and Y143 (phosphorylation). Of these, only K148 is included in the GnomAD dataset as K148(140)R (Fig. 5B; Supplemental Material). Of note, K148(140)R has a CADD C score of 25, which met our criteria for a CoF candidate.

In addition to the highlighted variant residues above, D141(133)N, R145(137)C, and L46(38)P met our structural and sequence analysis criteria. Residues D141(133)N and R145(137)C both have very high

![Fig. 5](https://example.com/fig5.png) The R12 subfamily of RGS proteins: analysis of RGS10 human variants. (A) Clustal W was used to align sequences of the RGS domains within the R12 subfamily of human RGS proteins. The RGS10-Ga\textit{i3-AlF4} crystal structure (Protein Data Bank: 2IHB) was used to extract residues on RGS10 critical for interaction with Ga\textit{i3} (highlighted in yellow). (B) MTR and PTM cluster analysis plot for RGS10-3 (UniProt ID O43665-3). The RGS domain of RGS10 (highlighted in blue) is under selective pressure, indicated by a low MTR (median, red line; 25%, gray dashed line; 5%, green dashed line). Human RGS10 contains many reported post-translational modifications, including multiple ubiquitinations, acetylations, and phosphorylations, one of which is found to be mutated in the RGS domain (GnomAD version 2.0) and met our criteria for predicted CoF phenotype (Supplemental Table 1; Table 2). (C) Rare human variants of RGS10 plotted along the sequence using Lollipops (https://github.com/pbnjay/lollipops). Missense (amino acid change) variants are displayed on top in teal, whereas silent (no amino acid change) variants are displayed on the bottom in purple. (D) From our analyses, we selected six variants in the RGS domain with predicted CoF phenotype and mapped them onto the sequence of RGS10 using Lollipops (https://github.com/pbnjay/lollipops). (E) Finally, we mapped these selected variants onto the crystal structure (Protein Data Bank: 2IHB) of RGS10 (gray, selected residues in red) and Ga\textit{i3-AlF4} (light blue).
CADD C scores of 35 and are close to two reported PTMs, a phosphorylation on Y143(135) and a ubiquitination on K148(140), indicating that these variants are likely to be deleterious to protein function. Consistent with this idea, both variants occur at highly conserved residues that are important for ionic stabilization. Changing the charge of the side chain from negative to uncharged, as is the case for D141(133)N and R145(137)C, is likely to disrupt efficient binding of the RGS domain to the Go protein. L46(38)P (C score 29.6) inserts a proline into an a helix, which we predict will cause disruption of the secondary structure. Furthermore, L46(38)P falls near two neighboring PTMs, a phosphorylation on S41(33) and a ubiquitination on K53(45). We postulate that disrupting secondary structure via insertion of a proline into an a helix would greatly inhibit the ability of a modifying enzyme to dock.

Finally, V52(44)M has a high CADD C score of 34 and is next to a modified residue, K53(45), suggesting that this mutation could lead to LoF. In support of this evidence, we observed that V52(44)M is putatively linked to schizophrenia (Hishimoto et al., 2004). V52(44) appears to participate in a hydrophobic core of the RGS domain; thus, a mutational change to methionine will likely disrupt this interaction. Interestingly, V52(44)M is found in nearly 2% of the East Asian population, and as such, we speculate that it could be an attractive target for deeper study. Furthermore, as loss of RGS10 protein in mice is linked with an increase in neuroinflammation (Lee et al., 2008, 2011, 2016; Lee and Tansey, 2015), LoF variants highlighted in this work may lead to susceptibility for neurodegenerative diseases such as Parkinson’s disease. Table 2 lists human variants in RGS10 that are predicted to disrupt RGS10 interaction with Go and/or GAP activity.

D. RZ Family: RGS17 Rare Variants

Unlike the other RGS subfamilies, no member of the RZ subfamily (RGS17, RGS19, or RGS20) has been crystalized in complex with a Go partner. However, a crystal structure of RGS17 alone has been reported (Soundararajan et al., 2008), allowing us to use this structure and conserved sequence alignments with other RGS family members to determine predicted CoF variants. RGS17 also has been linked to a number of diseases (Mao et al., 2004; James et al., 2009; You et al., 2009; Hayes and Roman, 2016), and therefore provides an attractive case study example for the RZ subfamily. For this, we first aligned the RGS domain of RGS17 with each of the case study proteins (RGS4, RGS9, RGS10; Figs. 2, 4, and 5) and highlight in yellow the residues that are indicated as contact points with Go, based on their respective crystal structures (Supplemental Fig. 1). We noted a great deal of conservation for residues that participate in Go interaction. For example, residues 145–155 correspond to the a5-a6 loop on each of the comparison family members (Tesmer et al., 1997; Slep et al., 2001; Soundararajan et al., 2008), a region found within RGS domains that is critical for stabilization of switch II in Go and GAP activity. We then used this information to predict contact points between RGS17 and Go, and highlighted in yellow those predicted contact points in Fig. 6A for all members of the RZ family. Each of those predicted contact points is either highly or completely conserved (highly conserved denoted by “,” completely conserved denoted by “*,” yellow boxes). Again, MTR analysis for RGS17 (Fig. 6B) indicates that the RGS domain is under selective pressure, with the exception of residues 134–143, which immediately precede the highly conserved a5-a6 loop. Using PTM alignment analysis, we identified two phosphorylation sites reported in the RGS domain of human RGS17, Y137 and Y171. Whether either phosphosite is functional for the protein is yet unknown. Although Y171 is a known variant site (Y171F), its low PTM observation frequency and CADD C score (9.1) exclude this variant from our selected dataset. Amino acid Y137 contributes to a MAP that is one of the most frequently observed phosphorylation sites in the RGS domain, but whose function is currently unknown. Thus, although not aligned with a genetic variant, PTM at this position appears to be very common in the RGS domain of several RGS proteins, a highly predictive feature for functional significance (Dewhurst et al., 2015; Torres et al., 2016; Dewhurst and Torres, 2017). Indeed, the lack of genetic variance at or near this position suggests that it is of critical importance, and any future variant at this position would be a prime target for functional analysis.

Based on our combined structural, conservation, CADD, and PTM analysis, three variants met our predictive criteria: E148G, P166L, and R189T (Fig. 6D). Amino acid E148 in RGS17 is a charged residue found in the highly conserved a5-a6 loop that serves as a key Go contact point for each of the representative RGS (Fig. 6A; Supplemental Fig. 1). E148 in RGS17 is also 100% conserved within the RZ family. The charged residues analogous to E148 in other RGS protein crystal structures participate in ionic bonding with opposite charge residues to stabilize this a5-a6 loop; thus, this negatively charged glutamate in RGS17 is a likely ionic binding partner serving the same role. As such, the E148G variant (C score 28.2) of RGS17 could disrupt RGS17 interactions with target Go subunits. We also identified P166L (C score 28.3), which is 100% conserved both within the RZ family (Fig. 6A) and across the other RGS subfamilies (Supplemental Fig. 1). This preserved proline appears to mediate the critical turn in the a6-a7 loop that is present in all RGS domains. Furthermore, it is located near a neighboring PTM, a phosphorylation on Y171. Due to the drastically different side chain properties of leucine versus proline, we speculate that P166L could result in CoF. Residue R189T (C score 28.1), which was also identified as a potential CoF candidate, like P166, is 100% conserved within the RZ subfamily and across all other RGS
We predict this variant could disrupt RGS17 ionic interactions with target Gα subunits, as shown in a structural homology model based on the solved crystal structure of the RGS4:Gαi1-AlF42 complex (Tesmer et al., 1997) (Fig. 6E).

IV. Pharmacological Impact of Human Variants in RGS Proteins

GPCRs have dominated the field of drug discovery for decades, and resulting new therapeutics have revolutionized modern medicine (Lundstrom, 2009). However, despite their success in ameliorating many diseases, drugs that target GPCRs often have off-target effects, making their therapeutic utility less than ideal. In more recent years, drug discovery efforts have turned to new ways of modulating GPCR signaling, including but not limited to such novel approaches as biased agonism (Bologna et al., 2017), positive and negative allosteric modulators (i.e., PAMs and NAMs) (Kenakin and Miller, 2010), and targeting the regulation of G protein signaling via RGS proteins (Sjögren et al., 2010). The rationale (Neubig and Siderovski, 2002) for targeting RGS proteins would be to do the following: 1) potentiate the effects of “dirty” GPCR agonists, thereby lowering the requisite therapeutic dose and off-target actions, and 2) enhance specific tissue and/or receptor signaling output when confronted with reduced natural agonist (e.g., depression, neurodegenerative diseases, others).

Currently, available small-molecule modulators that target RGS proteins do so by binding to and blocking RGS domain interactions with Gα (Hayes et al., 2018). The emergence of personal genomics raises the realistic prospect of personalized medicine (Cardon and Harris, 2016). Understanding the specific drug and target mechanism of action, and how that is intertwined with genetic variation (pharmacogenetics), could lead to better predictions of therapeutic efficacy.

The nascent field of pharmacogenetics aims to develop new medications that are tailored to an individual’s genetic makeup. As we move closer to this new world, the benefits of understanding the effects of genetic variation on target protein GoF versus LoF would be of great value, as demonstrated by recent reports of the profound impact of human variants on prescription drug effects on GPCR (Hauser et al., 2018) and NMDAR actions (Ogden et al., 2017). If a RGS protein is implicated in a complex disease, particularly by enhancement of the RGS domain/GAP activity, a RGS inhibitor may prove to be a superior target compared...
with a drug that targets a GPCR upstream of the RGS. For example, RGS5 mediates angiogenesis, and, remarkably, vascularization of tumors is normalized in RGS5 knockout mice (Hamzah et al., 2008). Thus, normal or GoF variants of RGS5 would promote tumor angiogenesis, and, in this case, a selective targeted inhibitor of RGS5 may make immune-mediated destruction of solid tumors more effective, greatly improving survival. Similarly, if a GoF mutation in the RGS domain of RGS4, for instance (Hill et al., 2008), is found to underlie an individual’s cardiovascular disease, a RGS inhibitor that selectively blocks RGS4 actions such as CCG-203769 (Blazer et al., 2015) may offer therapeutic benefit. By contrast, if a LoF in RGS4 or other RGS protein is found to contribute to disease progression, such a RGS inhibitor may exacerbate the disease and should thus be avoided. A prime example of the power of this analysis is a recent report on RGS2 human variants (Phan et al., 2015). The authors found that these RGS2 human variants, located in the N terminus, reduced RGS2-mediated inhibition of Ca\(^{2+}\) signaling via protein degradation and mislocalization and provided a mechanism for their association with hypertension. However, simply knowing a link between a variant and a disease does not account for functional directionality. Instead, one must determine whether the variant enhances or reduces protein function. Knowing this information becomes important in the case of RGS proteins because they regulate the timing of G protein on/off rates; thus, a drug that stimulates a GPCR may exacerbate the disease state in a patient with a LoF or GoF variant in a downstream regulatory RGS. Just as mutations in a metabolic enzyme dysregulate intended bioavailability or half-life of a drug, dysregulated RGS proteins may lead to extended or exaggerated drug actions. Therefore, investigators and clinicians will want to understand how natural genetic variation may affect druggable targets, both current and future.

Although we have focused this review on the effects of rare human variants within the canonical RGS domain, we must note that other domains and regions on RGS proteins are, in many ways, equally important for RGS protein function. Other RGS protein domains/regions dictate subcellular localization, modify GAP activity, stabilize the protein, and engage additional signaling pathways independent of G protein signaling. For example, RGS4 is robustly degraded by an N-terminal cysteine degradation signal (Bodenstein et al., 2007), and genetic modification (C→S) of this N-terminal cysteine increases RGS4 protein expression by nearly 50-fold. This could have major implications in diseases linked with RGS4 expression, such as schizophrenia and alcoholism (Mirmics et al., 2001; Ho et al., 2010). The subcellular localization of RGS14 is regulated by a C-terminal GPR motif (Shu et al., 2007; Brown et al., 2015b), which binds an inactive Ga-GDP. Ablation of this interaction may have consequences for the role of RGS14 as a regulator of hippocampal-based learning and synaptic long-term potentiation (Lee et al., 2010). RGS7 family members all contain DEP domains (binds R7BP) that are important for plasma membrane localization (Drenan et al., 2006) and GGL domains (binds Gβ5) that are essential for protein stability (Chen et al., 2003). Loss of interaction with either of these critical binding partners would disrupt the function of RGS7 completely independent of the RGS domain, and the DEP domain could, in theory, serve as a beneficial drug target. Indeed, information gained from LoF or GoF genetic variation throughout each of the functional domains of these heterogeneous RGS proteins may provide key insight into the etiology of disease and, should drug modulators become available, may offer the best course of treatment. The widespread accessibility of genome sequencing and parallel improvements in in silico predictions of protein functions brings precision medicine and pharmacogenomics much closer to reality. Within this context, modern drug development and usage should take into account whether, and how, natural genetic variation may affect patient responses to therapies targeting complex disease states.

Acknowledgments
We thank Joshua L. Traynelis, who kindly provided MTR analysis and data and constructive suggestions on how to improve the manuscript. We thank Suneela Ramineni for technical support and Kyle Gerber for motivation to work faster. We also thank Jeff Squires, who provided endless guidance on using Excel for complex data management.

Authorship Contributions
Contributed experiments: Squires, Montañez-Miranda, Hepler. Contributed new reagents or analytic tools: Squires, Pandya, Torres.
Wrote or contributed to the writing of the manuscript: Squires, Montañez-Miranda, Pandya, Torres, Hepler.

References
Bergers G and Song S (2005) The role of pericytes in blood-vessel formation and
Barker SA, Wang J, Sierra DA, and Ross EM (2001) RGSZ1 and Ret RGS: two of
Bansal G, Xie Z, Rao S, Nocka KH, and Druey KM (2008b) Suppression of immu-
Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E,
Bodenstein J, Sunahara RK, and Neubig RR (2007) N-terminal residues control
468 Squires et al.
Cardon LR and Harris T (2016) Precision medicine, genomics and drug discovery.
Ham Mol Genet 58:166–172.
Chen JR, Sedding DG (2016) Regulator of G-protein signaling 5 prevents smooth muscle

Bansal G, Xie Z, Rao S, Nocka KH, and Druey KM (2008b) Suppression of immu-


Genetic Variation in Human RGS Proteins


Snow BE, Krumins AM, Brothers GM, Lee SF, Wall MA, Chung S, Mangion J, Arya
Tansey MG and Goldberg MS (2010) Neuroinflammation in Parkinson
Tennessen JA, Bigham AW, O
Sjögren B, Blazer LL, and Neubig RR (2010) Regulators of G protein signalling pro-
Torres
Tso PH, Lyon YG, Wang Y, and Wong YH (2011) RGS91 stimulates cell proliferation by deregulating cell cycle control and enhancing Akt signaling. Cancer Lett 309:
Wu Q and Traynor JR (2013) 6462.


Supplemental Figures, Tables and Data

Methods

Human RGS protein sequence references (UniProt IDs)

For data shown in Figures and Tables, the following human RGS protein exome sequences were analyzed and presented. The UniProt IDs for each of these are: RGS1: Q08116-1; RGS2: P41220-1; RGS3: P49796-3; RGS4: P49798-3 (longest RGS4-3) and P49798-1 (canonical RGS4-1); RGS5: O15539-3; RGS6: P49758-3; RGS7: P49802-1; RGS8: P57771-2; RGS9: O75916-1; RGS10: O43665-3 (longest RGS10-3) and O43665-1 (canonical RGS10-1); RGS11: O94810-1; RGS12: O14924-1; RGS13: O14921-1; RGS14: O43566-7; RGS16: O15492-1; RGS17: Q9UGC6-1; RGS18: Q9NS28-1; RGS19: P49795-1; RGS20: O76081-1; RGS21: Q2M5E4-1.

Immunoaffinity capture of the RGS4:Gαi1 complex from cells

HEK293 cells were transfected with Venus-RGS4 (WT or R134W) and Gαi1-Luciferase. Cells were washed with ice cold PBS, and then lysed with AlF₄⁻ lysis buffer (50mM Tris, 150mM NaCl, 1mM EDTA, 2mM DTT, 5mM MgCl₂, 1% Triton X-100, protease inhibitors, 10mM NaF, 9mM MgCl₂, 30µM AlCl₃). Cells rotated end-over-end for 1.5 hours at 4°C. Samples were then centrifuged for 10 minutes at 13,000 RPM at 4°C, and 30µL of cell lysate was collected for input. Protein G sepharose (GE Healthcare) (50μL) was washed 3x with 1mL of ice cold PBS, and then incubated for an hour with 3% BSA (bovine serum albumin) rotating at 4°C. Beads were spun down, 3% BSA was removed, and lysates were added to beads with 2µL anti-GFP (MBL) to rotate for 1.5 hours at 4°C. Beads were then centrifuged for 30 seconds at 3000 RPM at 4°C, washed 4x with 1mL of ice cold wash buffer (0.1% Tween-20 in 1x PBS with AlF₄⁻), and boiled in Laemmli buffer for 5 minutes. 13% SDS-PAGE separated samples and nitrocellulose membranes were blotted with anti-Renilla Luciferase (Millipore, 1:1000) and anti-GFP (MBL, 1:1000). Secondary antibodies used were goat anti-mouse (1:5000) and goat anti-rabbit (1:25000), respectively.
**Bioluminescence Resonance Energy Transfer (BRET)**

For variant and wild type RGS4, we generated RGS4-Luciferase (RGS4-Luc) fusion constructs, a tag suitable for bioluminescence resonance energy transmission (BRET) which emits light at 485nm in the presence of 5 μM coelenterazine H (CTZ) (Brown et al., 2015a). In combination with a YFP-tagged Goi1 (Goα-YFP), RGS4-Luc will transfer energy to Goα-YFP upon binding, thereby producing a 535nm emission BRET signal. We incubated AlF4- with cells expressing RGS4-Luc and Goα-YFP to mimic the transition state of Goα and recruit RGS4 binding for 30 minutes and then recorded emissions at 485nm and 535nm following CTZ activation. Net BRET was calculated by dividing the emission at 535nm by the emission at 485nm, and then subtracting background signal of 485nm alone.

**Missense Tolerance Ratio**

We calculated missense tolerance ratio as recently described (Traynelis et al., 2017). Briefly, we used a 31-codon sliding window using the following calculation:

\[
MTR = \frac{\sum[\text{missense}_{obs}] / \sum[\text{missense}_{obs} + \text{synonymous}_{obs}]}{\sum[\text{missense}_{expected}] / \sum[\text{missense}_{expected} + \text{synonymous}_{expected}]}
\]

where observed variants ("obs") were reported variants from the GnomAD database and expected variants were calculated from all possible codon variations.

**Post-Translational Modification Alignment Analysis**

PTM alignment analysis is based on the SAPH-ire model as described previously (Torres et al., 2016) but was restricted to report only PTM count as described below. All PTMs are experimentally verified and were retrieved from dbPTM and SysPTM (Huang et al., 2016b; Li et al., 2014). To account for the diverse domains observed across the different proteins of the RGS protein family, a domain specific analysis was performed for all domain families observed within the RGS superfamily. Seven domains seen across RGS proteins were analyzed, namely DEP: disheveled, EGL-10, pleckstrin homology domain; DHEX, DEP helical extension (IPR000591), GGL: Gγ subunit-like domain (IPR015898), GPR/GoLoco: G protein regulatory motif (IPR003109), PDZ: domain present in PSD-95, Dlg, and ZO-1/2 (IPR001478), PTB: phospho-tyrosine binding domain (IPR006020), RBD: Raf-like Ras binding domain (IPR003116), and
The RGS domain (IPR016137; the focus of this review). The domain sequences for all RGS proteins belonging to each of these domain families were obtained from Interpro (Hunter et al., 2009). The multi-FASTA files generated for each domain family were subsequently aligned using MUSCLE using default parameters (Edgar, 2004). Modified Alignment Positions (MAPs), domain family alignment positions that harbor at least one experimentally verified PTM specific to human RGS proteins, were coalesced and the PTM count within each determined by the sum of human family members for which experimental PTM has been observed and curated in the public domain via dbPTM and Phosphosite Plus (Hornbeck et al., 2015; Huang et al., 2016b). PTMs outside the RGS domain and specific to each given RGS protein are reported as single observations though may, in many cases, coalesce into MAPs of other domain families (data not shown). Non-domain PTMs and RGS domain MAPs were related directly to mutation data based on UniProt ID and native position anchor sequences (Bairoch et al., 2005), and plotted with respect to the MTR value at each native position.

SUPPLEMENTAL DATA

Table S1. RGS protein variants that overlap with post-translational modifications (PTMs). Post translational modifications that overlap with a human variant (GnomAD) for each RGS1-21 are listed. Grey boxes indicate the PTM/Mutation overlap is found within the canonical RGS domain. The type of modification, missense tolerance ratio (MTR), combined annotation dependent depletion (CADD) C-score, and population allele frequency (Prevalence) are listed. Allele frequency is listed in terms of the most prevalent population. If multiple populations have high prevalence, then prevalence in listed in terms of global allele frequency. Human variants are described according to GnomAD convention based on the longest splice variant for each human RGS protein, as opposed to the canonical (Uniprot-based) isoform. Thus, we include in parenthesis the canonical residue number in cases where they differ (RGS4, RGS5 and RGS10) from the longest splice variant position.
<table>
<thead>
<tr>
<th>RGS</th>
<th>Residue Mutation</th>
<th>PTM</th>
<th>CADD</th>
<th>MTR</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGS1</td>
<td>K13E</td>
<td>Methylation</td>
<td>16.31</td>
<td>0.99</td>
<td>0.0065% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>S67Y</td>
<td>Phosphorylation</td>
<td>22.7</td>
<td>0.92</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>K164Q</td>
<td>Methylation</td>
<td>7.4</td>
<td>0.95</td>
<td>0.0032% (South Asian)</td>
</tr>
<tr>
<td>RGS2</td>
<td>S46N</td>
<td>Phosphorylation</td>
<td>22.5</td>
<td>0.70</td>
<td>0.00232% (East Asian)</td>
</tr>
<tr>
<td></td>
<td>S64G</td>
<td>Phosphorylation</td>
<td>13.77</td>
<td>0.92</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>Y116C</td>
<td>Phosphorylation</td>
<td>25.5</td>
<td>0.92</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>Y133C</td>
<td>Phosphorylation</td>
<td>27</td>
<td>0.88</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>Y178C</td>
<td>Phosphorylation</td>
<td>NA</td>
<td>0.90</td>
<td>0.0182% (Other)</td>
</tr>
<tr>
<td>RGS3</td>
<td>S395L</td>
<td>Phosphorylation</td>
<td>29.3</td>
<td>0.97</td>
<td>0.0037% (Global)</td>
</tr>
<tr>
<td></td>
<td>T396I</td>
<td>Phosphorylation</td>
<td>23.3</td>
<td>0.99</td>
<td>0.0033% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>R448H</td>
<td>Methylation</td>
<td>14.5</td>
<td>0.73</td>
<td>0.0210% (Other)</td>
</tr>
<tr>
<td></td>
<td>S496G</td>
<td>Phosphorylation</td>
<td>14.3</td>
<td>1.09</td>
<td>0.0060% (Latino)</td>
</tr>
<tr>
<td></td>
<td>S496N</td>
<td>Phosphorylation</td>
<td>12.11</td>
<td>1.09</td>
<td>0.0065% (African)</td>
</tr>
<tr>
<td></td>
<td>T536M</td>
<td>Phosphorylation</td>
<td>23.9</td>
<td>0.67</td>
<td>0.0089% (Latino)</td>
</tr>
<tr>
<td></td>
<td>S662L</td>
<td>Phosphorylation</td>
<td>27</td>
<td>0.63</td>
<td>0.0455% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>S777L</td>
<td>Phosphorylation</td>
<td>18.2</td>
<td>0.81</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>S806F</td>
<td>Phosphorylation</td>
<td>19.4</td>
<td>0.82</td>
<td>0.0060% (Latino)</td>
</tr>
<tr>
<td></td>
<td>T808N</td>
<td>Phosphorylation</td>
<td>15.6</td>
<td>0.84</td>
<td>0.1674% (Global)</td>
</tr>
<tr>
<td></td>
<td>S823F</td>
<td>Phosphorylation</td>
<td>14.8</td>
<td>1.07</td>
<td>0.0018% (European)</td>
</tr>
<tr>
<td></td>
<td>T921M</td>
<td>Phosphorylation</td>
<td>28.8</td>
<td>0.95</td>
<td>0.0390% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>S936L</td>
<td>Phosphorylation</td>
<td>34</td>
<td>0.87</td>
<td>0.1046% (African)</td>
</tr>
<tr>
<td></td>
<td>T941M</td>
<td>Phosphorylation</td>
<td>34</td>
<td>0.90</td>
<td>0.0033% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>S965L</td>
<td>Phosphorylation</td>
<td>13.4</td>
<td>0.75</td>
<td>0.0033% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>S992C</td>
<td>Phosphorylation</td>
<td>25</td>
<td>1.16</td>
<td>0.0067% (European)</td>
</tr>
<tr>
<td></td>
<td>S1007N</td>
<td>Phosphorylation</td>
<td>22</td>
<td>0.88</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td>RGS4</td>
<td>C109(12)R</td>
<td>Palmitoylation</td>
<td>24.9</td>
<td>0.60</td>
<td>0.0083% (African)</td>
</tr>
<tr>
<td>RGS5</td>
<td>Y73C</td>
<td>Phosphorylation</td>
<td>17.8</td>
<td>0.88</td>
<td>0.0041% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>Y117C</td>
<td>Phosphorylation</td>
<td>26.5</td>
<td>0.94</td>
<td>0.0167% (African)</td>
</tr>
<tr>
<td></td>
<td>S170 (166) P</td>
<td>Phosphorylation</td>
<td>32</td>
<td>1.13</td>
<td>0.0117% (East Asian)</td>
</tr>
<tr>
<td>RGS7</td>
<td>K251Q</td>
<td>Acetylation</td>
<td>24.5</td>
<td>0.93</td>
<td>0.0058% (East Asian)</td>
</tr>
<tr>
<td></td>
<td>S457L</td>
<td>Phosphorylation</td>
<td>21.8</td>
<td>0.69</td>
<td>0.0058% (East Asian)</td>
</tr>
<tr>
<td>Protein</td>
<td>Site</td>
<td>Modification</td>
<td>Value</td>
<td>SD</td>
<td>Percentage</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>--------------</td>
<td>-------</td>
<td>----</td>
<td>------------</td>
</tr>
<tr>
<td>RGS9</td>
<td>Y153C</td>
<td>Phosphorylation</td>
<td>25.8</td>
<td>0.85</td>
<td>0.0155% (Other)</td>
</tr>
<tr>
<td></td>
<td>Y153S</td>
<td>Phosphorylation</td>
<td>23.5</td>
<td>0.85</td>
<td>0.0030% (Latino)</td>
</tr>
<tr>
<td></td>
<td>Y153D</td>
<td>Phosphorylation</td>
<td>23.4</td>
<td>0.85</td>
<td>0.0016% (European)</td>
</tr>
<tr>
<td></td>
<td>S304R</td>
<td>Phosphorylation</td>
<td>3.7</td>
<td>0.72</td>
<td>0.0016% (European)</td>
</tr>
<tr>
<td></td>
<td>Y413C</td>
<td>Phosphorylation</td>
<td>27.2</td>
<td>0.94</td>
<td>0.0032% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>K419R</td>
<td>Acetylation</td>
<td>17.12</td>
<td>0.93</td>
<td>0.0029% (Latino)</td>
</tr>
<tr>
<td></td>
<td>S24(16)T</td>
<td>Phosphorylation</td>
<td>7.27</td>
<td>0.78</td>
<td>0.0030% (Latino)</td>
</tr>
<tr>
<td>RGS10</td>
<td>S27(19)N</td>
<td>Phosphorylation</td>
<td>21.9</td>
<td>0.81</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>K148(140)R</td>
<td>Ubiquitination</td>
<td>25</td>
<td>0.81</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>Y179(171)N</td>
<td>Phosphorylation</td>
<td>28.5</td>
<td>1.1</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td>RGS12</td>
<td>S16L</td>
<td>Phosphorylation</td>
<td>11.27</td>
<td>0.91</td>
<td>0.0066% (African)</td>
</tr>
<tr>
<td></td>
<td>S16A</td>
<td>Phosphorylation</td>
<td>0.001</td>
<td>0.91</td>
<td>13.0309% (African)</td>
</tr>
<tr>
<td></td>
<td>S88P</td>
<td>Phosphorylation</td>
<td>23.1</td>
<td>0.93</td>
<td>0.0032% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>S107G</td>
<td>Phosphorylation</td>
<td>24</td>
<td>1.00</td>
<td>0.0032% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>S107C</td>
<td>Phosphorylation</td>
<td>23.8</td>
<td>1.00</td>
<td>0.0054% (European)</td>
</tr>
<tr>
<td></td>
<td>S172L</td>
<td>Phosphorylation</td>
<td>26.8</td>
<td>0.99</td>
<td>0.0155% (Other)</td>
</tr>
<tr>
<td></td>
<td>S195C</td>
<td>Phosphorylation</td>
<td>23.2</td>
<td>1.05</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>S284N</td>
<td>Phosphorylation</td>
<td>13.8</td>
<td>0.93</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>S483I</td>
<td>Phosphorylation</td>
<td>0.64</td>
<td>0.95</td>
<td>0.0010% (European)</td>
</tr>
<tr>
<td></td>
<td>T492N</td>
<td>Phosphorylation</td>
<td>0.001</td>
<td>0.97</td>
<td>0.0028% (European)</td>
</tr>
<tr>
<td></td>
<td>R633W</td>
<td>Methylation</td>
<td>31</td>
<td>1.09</td>
<td>0.0032% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>R633P</td>
<td>Methylation</td>
<td>23.8</td>
<td>1.09</td>
<td>0.0167% (African)</td>
</tr>
<tr>
<td></td>
<td>R633Q</td>
<td>Methylation</td>
<td>23.5</td>
<td>1.09</td>
<td>8.5678% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>R633L</td>
<td>Methylation</td>
<td>21.1</td>
<td>1.09</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>R643C</td>
<td>Methylation</td>
<td>29</td>
<td>1.03</td>
<td>0.0682% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>R643S</td>
<td>Methylation</td>
<td>27.8</td>
<td>1.03</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>R643H</td>
<td>Methylation</td>
<td>26.4</td>
<td>1.03</td>
<td>0.0365% (Other)</td>
</tr>
<tr>
<td></td>
<td>S649L</td>
<td>Phosphorylation</td>
<td>26.9</td>
<td>0.92</td>
<td>0.0065% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>S649P</td>
<td>Phosphorylation</td>
<td>23.3</td>
<td>0.92</td>
<td>0.0032% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>S649T</td>
<td>Phosphorylation</td>
<td>23.2</td>
<td>0.92</td>
<td>0.0030% (Latino)</td>
</tr>
<tr>
<td></td>
<td>T676M</td>
<td>Phosphorylation</td>
<td>25</td>
<td>0.75</td>
<td>0.1193% (Latino)</td>
</tr>
<tr>
<td></td>
<td>T866M</td>
<td>Phosphorylation</td>
<td>32</td>
<td>0.82</td>
<td>0.0054% (East Asian)</td>
</tr>
<tr>
<td></td>
<td>S934L</td>
<td>Phosphorylation</td>
<td>30</td>
<td>0.78</td>
<td>0.0058% (East Asian)</td>
</tr>
<tr>
<td></td>
<td>S947L</td>
<td>Phosphorylation</td>
<td>17.9</td>
<td>0.56</td>
<td>0.0104% (Ashkenazi)</td>
</tr>
<tr>
<td></td>
<td>T1331M</td>
<td>Phosphorylation</td>
<td>21.6</td>
<td>0.67</td>
<td>0.0092% (Global)</td>
</tr>
<tr>
<td>Gene</td>
<td>Position</td>
<td>Modification</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
</tr>
<tr>
<td>--------</td>
<td>----------</td>
<td>--------------</td>
<td>---------</td>
<td>---------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>RGS13</td>
<td>Y146H</td>
<td>Phosphorylation</td>
<td>27.3</td>
<td>0.99</td>
<td>0.0535% (East Asian)</td>
</tr>
<tr>
<td></td>
<td>R27M</td>
<td>Methylation</td>
<td>21.4</td>
<td>1.03</td>
<td>0.0191% (Other)</td>
</tr>
<tr>
<td></td>
<td>K86N</td>
<td>Ubiquitination</td>
<td>27.9</td>
<td>0.80</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>K86R</td>
<td>Ubiquitination</td>
<td>23.5</td>
<td>0.80</td>
<td>0.0030% (Latino)</td>
</tr>
<tr>
<td></td>
<td>S260F</td>
<td>Phosphorylation</td>
<td>32</td>
<td>0.97</td>
<td>0.0754% (East Asian)</td>
</tr>
<tr>
<td></td>
<td>S288R</td>
<td>Phosphorylation</td>
<td>29.1</td>
<td>1.11</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>T292K</td>
<td>Phosphorylation</td>
<td>12.15</td>
<td>1.12</td>
<td>0.1166% (East Asian)</td>
</tr>
<tr>
<td></td>
<td>T292M</td>
<td>Phosphorylation</td>
<td>13.01</td>
<td>1.12</td>
<td>0.0030% (Latino)</td>
</tr>
<tr>
<td></td>
<td>T472A</td>
<td>Phosphorylation</td>
<td>0.371</td>
<td>1.08</td>
<td>0.2729% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>S480G</td>
<td>Phosphorylation</td>
<td>6.092</td>
<td>0.98</td>
<td>0.0067% (European)</td>
</tr>
<tr>
<td></td>
<td>T490S</td>
<td>Phosphorylation</td>
<td>0.004</td>
<td>0.94</td>
<td>0.0060% (Latino)</td>
</tr>
<tr>
<td></td>
<td>T490N</td>
<td>Phosphorylation</td>
<td>12.99</td>
<td>0.94</td>
<td>0.0162% (South Asian)</td>
</tr>
<tr>
<td></td>
<td>S510G</td>
<td>Phosphorylation</td>
<td>23.4</td>
<td>1.06</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td>RGS14</td>
<td>C12</td>
<td>Palmitoylation</td>
<td>33</td>
<td>1.11</td>
<td>0.0115% (African)</td>
</tr>
<tr>
<td></td>
<td>C98Y</td>
<td>Palmitoylation</td>
<td>33</td>
<td>0.96</td>
<td>0.0047% (European)</td>
</tr>
<tr>
<td>RGS16</td>
<td>Y171F</td>
<td>Phosphorylation</td>
<td>9.1</td>
<td>0.77</td>
<td>0.0067% (European)</td>
</tr>
<tr>
<td>RGS17</td>
<td>Y146C</td>
<td>Phosphorylation</td>
<td>26.3</td>
<td>0.64</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>T186A</td>
<td>Phosphorylation</td>
<td>16.4</td>
<td>0.67</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>T201I</td>
<td>Phosphorylation</td>
<td>11.5</td>
<td>0.90</td>
<td>0.0511% (Ashkenazi Jewish)</td>
</tr>
<tr>
<td></td>
<td>S211L</td>
<td>Phosphorylation</td>
<td>22.8</td>
<td>1.06</td>
<td>0.0115% (African)</td>
</tr>
<tr>
<td></td>
<td>S213A</td>
<td>Phosphorylation</td>
<td>0.11</td>
<td>1.07</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td>RGS19</td>
<td>S77R</td>
<td>Phosphorylation</td>
<td>23.9</td>
<td>1.13</td>
<td>0.0119% (Latino)</td>
</tr>
<tr>
<td></td>
<td>S77T</td>
<td>Phosphorylation</td>
<td>23</td>
<td>1.13</td>
<td>0.0072% (European)</td>
</tr>
<tr>
<td></td>
<td>T162I</td>
<td>Phosphorylation</td>
<td>15.7</td>
<td>0.90</td>
<td>0.0067% (European)</td>
</tr>
<tr>
<td></td>
<td>S166G</td>
<td>Phosphorylation</td>
<td>13.8</td>
<td>1.15</td>
<td>0.0067% (European)</td>
</tr>
<tr>
<td>RGS20</td>
<td>Y116C</td>
<td>Phosphorylation</td>
<td>25.5</td>
<td>0.92</td>
<td>0.0009% (European)</td>
</tr>
<tr>
<td></td>
<td>Y133C</td>
<td>Phosphorylation</td>
<td>27</td>
<td>0.88</td>
<td>0.0009% (European)</td>
</tr>
</tbody>
</table>
Supplemental Figures

**RGS Domain – Family Comparison**

Supplemental Raw Data Files

We include MTR plots with overlayed PTMs for the remaining RGS proteins that were not highlighted in the body of the text. We also include our comprehensive curated RGS PTM/Mutation database with all reported human variants for all canonical human RGS proteins from the GnomAD consortium (version 2.0). These are available as downloadable and expandable MS Excel Files (XLSX). We include only variants that were not filtered out by the Broad Institute to ensure quality source data. Human variants are described according to GnomAD convention based on the longest splice variant for each human RGS protein, as opposed to the canonical (UniProt-based) reference isoform. Thus, we include in parenthesis the canonical residue number in cases where they differ from the longest splice variant position.

Figure S1. Family alignment of RGS domains. To identify conserved residues critical for Gα interaction, we used Clustal W to align each of the representative subfamily members (RGS4, RGS9, RGS10) with RGS17. We highlighted residues that were noted as important for binding (Figures 2, 4, 5) to compare against RGS17. From this information, we identified residues in RGS17 that are predicted contact points with Gα.
Figure S2. Analysis of rare variants and PTMs found within human RGS proteins. Data is presented for all human RGS proteins that were not highlighted in the body of the text. Human variants correspond to GnomAD convention based on the longest splice variant for each human RGS protein (see Methods for UniProt IDs).  

(For each RGS, TOP) Rare human variants of RGS proteins are plotted along the protein sequence using Lollipops (https://github.com/pbnjay/lollipops). Missense (amino acid change) variants are displayed on top in teal, while silent (no amino acid change) variants are displayed on the bottom in purple.  

(For each RGS, BOTTOM) Missense Tolerance Ratio (MTR) and Post-Translational Modification (PTM) cluster analysis plot for each RGS (see text for details). If the RGS domain (highlighted in blue) or other protein region is under selective pressure, the MTR will fall below the median (red line). The location of functional or disease-linked post-translational modifications (PTMs) are plotted as colored circles on the line. The PTM identities and amino acid location, as well as the MTR data for each RGS are available in Table S1 and in the downloadable and expandable MS Excel Files (XLSX).
RGS1

Missense

Disease-linked mutation site

Functional Family PTM

Functional Protein PTM

MTR

Obs. Freq.

RGS1

1 5 9

Native Position

0 20 40 60 80 100 120 140 160 180 200

RGS2

Missense

Disease-linked mutation site

Functional Family PTM

Functional Protein PTM

MTR

Obs. Freq.

RGS2

1 5 9

Native Position

0 20 40 60 80 100 120 140 160 180 200 220
**RGS3**

- Missense
- Silent

**RGS5**

- Missense
- Silent

**Functional Family PTM**

- Disease-linked mutation site

**Functional Protein PTM**

**Native Position**

- MTR

**Obs. Freq.**

- 1
- 5
- 9
RGS8

Missense
Silent

RGS8

Native Position

Disease-linked mutation site

Functional Family PTM

Functional Protein PTM

Obs. Freq.

1
5
9

Native Position

MTR

RGS8

0
20
40
60
80
100
120
140
160
180
200

0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

Obs. Freq.

1
5
9

Native Position

MTR

RGS11

Missense
Silent

RGS11

Native Position

Disease-linked mutation site

Functional Family PTM

Functional Protein PTM

Obs. Freq.

1
5
9

Native Position

MTR

RGS11

0
40
80
120
160
200
240
280
320
360
400
440
480

0
0.2
0.4
0.6
0.8
1.0
1.2
1.4

Obs. Freq.

1
5
9

Native Position

MTR

RGS11

0
40
80
120
160
200
240
280
320
360
400
440
480

0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
RGS19

Missense

Silent

RGS20

Missense

Silent

Functional Family PTM

Disease-linked mutation site

Functional Protein PTM

Native Position

MTR

Obs. Freq.

0 20 40 60 80 100 120 140 160 180 200 220

0 0.2 0.4 0.6 0.8 1.0 1.2 1.4

0 40 80 120 160 200 240 280 320 360 400

0 0.2 0.4 0.6 0.8 1.0 1.2 1.4

RGS19

RGS20
Obs. Freq.
1
9

Functional Family PTM
Disease-linked mutation site

Functional Protein PTM

Native Position

RGS21

Missense
Silent

MTR

0 20 40 60 80 100 120 140 160

RGS21

Obs. Freq.
5

1

Disease-linked mutation site

Functional Family PTM
Functional Protein PTM

Native Position

RGS21