Gαs-PKA pathway signalopathies: The emerging genetic landscape and therapeutic potential of human diseases driven by aberrant Gαs-PKA signaling

Dana J. Ramms, Francesco Raimondi, Nadia Arang, Friedrich W. Herberg, Susan S. Taylor, and J. Silvio Gutkind

Pharmacological Reviews

PHARMREV-AR-2020-000269R1

Supplementary Files

Table S1. Expression of Gαs-PKA pathway components

Table S2. Pathway mutational data from genetic diseases

Table S3. Residue mutational data in cancer

Table S4. Cancer fusion protein data

Table S5. Pathway mutational data in cancer

Supplementary Methods

Normal gene expression analysis

Median gene-level expression counts for normal tissue were download from the GTex Portal [https://www.gtexportal.org/home/datasets] (v8, June 2017). Results were curated for pathway genes provided in Table S1. For organs with multiple data entries (i.e. Brain - Amygdala, Brain - Anterior cingulate cortex…) values were averaged to represent whole organ gene expression. All values were represented in transcripts per million (TPM).

Variant annotation and interpretation

All available variant coordinates were retrieved from the ClinVar database [https://www.ncbi.nlm.nih.gov/pubmed/31777943/] as of February 2021. Variants were called using the human genome version GRCh37 as a reference [https://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh37/]. Variants were then annotated on Uniprot canonical sequences through Variant Effect Predictor [https://pubmed.ncbi.nlm.nih.gov/27268795/]. The ClinVar variant summary file [https://ftp.ncbi.nlm.nih.gov/pub/ clinvar/tab_delimited/] was used to retrieve variant curations, including “ClinicalSignificance”, which classifies variants as either pathogenic (or likely pathogenic), risk factors, benign or variants of unknown significance based on supporting evidence. Disease phenotype associations were retrieved using the information provided in the ”PhenotypeList” classification. An interaction network was obtained by querying the Reactome Functional Interaction (FI) network [https://pubmed.ncbi.nlm.nih.gov/28150241/] and the STRING database [https://pubmed.ncbi.nlm.nih.gov/30476243/] with the list of genes involved in the PKA pathway (Table S2). Networks were implemented using the Cytoscape apps Reactome FIVIZ [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4184317/] and stringApp

[https://pubmed.ncbi.nlm.nih.gov/30450911/] (confidence score > 0.75). The diameter of the nodes in the network were scaled proportionally to the number of variants classified as pathogenic, likely pathogenic, or risk factor. Pie charts with PhenotypeList classifications, when available, were added to network nodes through enhanced Graphics Cytoscape app [https://pubmed.ncbi.nlm.nih.gov/25285206/]. Phenotype labels for the pie chart representation were given through the following criteria: if "Acrodysostosis" matched PhenotypeList, label="Acrodysostosis"; if "Long QT syndrome" matched PhenotypeList, label="Long QT syndrome"; if "pseudohypoparathyroidism" or "PSEUDOHYPOPARATHYROIDISM" matched PhenotypeList, label="Pseudohypoparathyroidism"; if "McCune-Albright syndrome" matched PhenotypeList, label="McCune-Albright syndrome"; if "Cardiac" or "Cardio" matched PhenotypeList , label="Cardiovascular phenotype"; if "PITUITARY TUMOR 3" or "Pituitary adenoma 3" matched PhenotypeList, label="Pituitary adenoma"; if "CARDIOACROFACIAL DYSPLASIA 1" or "CARDIOACROFACIAL DYSPLASIA 2" matched PhenotypeList, label="CARDIOACROFACIAL DYSPLASIA"; if "Thrombocythemia" matched PhenotypeList, label="Thrombocythemia"; if "Dyskinesia" or "Dystonia" or "Striatal degeneration" matched PhenotypeList, label="Movement disorders".

Pathogenic mutations were annotated on protein canonical sequence diagrams through the Lollipops software [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4973895/]. Pathogenic mutations were then mapped on the 3D structure whenever available, alternatively we displayed any mutation available for a given interface of interest (i.e. AKAP10). Protein Data Bank (PDB) identifiers for structures displayed are as follows: β_2AR-Gαs (PDB: 3SN6), adenylyl cyclase-Gαs (PDB: 1AZS), PKA RIα-Cα (PDB: 5JR7), and AKAP10-RIα (PDB: 3IM4). Prediction of functional consequences of mutations at 3D interaction interfaces was then performed through Mechismo [https://pubmed.ncbi.nlm.nih.gov/25392414/] using default parameters.

Fusion transcript analysis

Fusion events were retrieved from the Fusion GDB database [https://pubmed.ncbi.nlm.nih.gov/30407583/]. Results were curated using the gene list provided in Table S4. The number of fusions identified for each gene was counted across cancer tissue types. Cancer tissue type data was pooled based on organ (i.e. “Lung” represents lung adenocarcinoma and lung squamous cell carcinoma). Drawing of fusion event statistic for the *GNAS* gene model were obtained through the ensembldb [https://pubmed.ncbi.nlm.nih.gov/ 30689724/] and gviz [https://www.springer.com/ gp/book/9781493935765] R libraries.

Analysis of cancer mutations

For analysis of residue mutation frequency, gene specific data was downloaded directly from the gene variant section of the COSMIC database in February 2021, COSMIC v92. Mutations were called with respect to the reference genome GRCh38 and all available mutations were included. Direct access links are provided below.

For *GNAS*: [https://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=GNAS#variants].

For *PRKACA*: [https://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=PRKACA#variants].

For *PRKACB*: [https://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=PRKACB#variants].

For *PRKAR1A*: [https://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=PRKAR1A#variants]. Mutation counts were pooled by residue and converted to frequencies. Only residue frequencies greater than 1% for their respective gene were considered recurrent for plotting purposes (Table S3).

For gene mutation analysis, whole genome screen data was downloaded directly from the COSMIC database [https://cancer.sanger.ac.uk/cosmic/download] in May 2020, COSMIC v91.

Mutation frequencies were curated across all available tissues. The corresponding “HGNC_ID” (Table S5) was used to extract gene-level mutation data. The data was then collapsed using the corresponding “ID_Sample” to ensure that no given sample was counted more than once. Gene family-level mutation frequencies were generated by counting unique patient samples harboring at least one mutation in a family gene member (family gene members listed in Table S5). The total number of individual pathway mutations per patient sample was also counted across tissues using “ID_Sample” and the “Pathway Family” gene set. All available mutations were considered for each analysis.