



# Pegasus Users Group

MEETING



# Automated Processing of Phenotypic Data Submissions using Pegasus

---

**Rajiv Mayani**

Research Programmer II

February 25, 2021

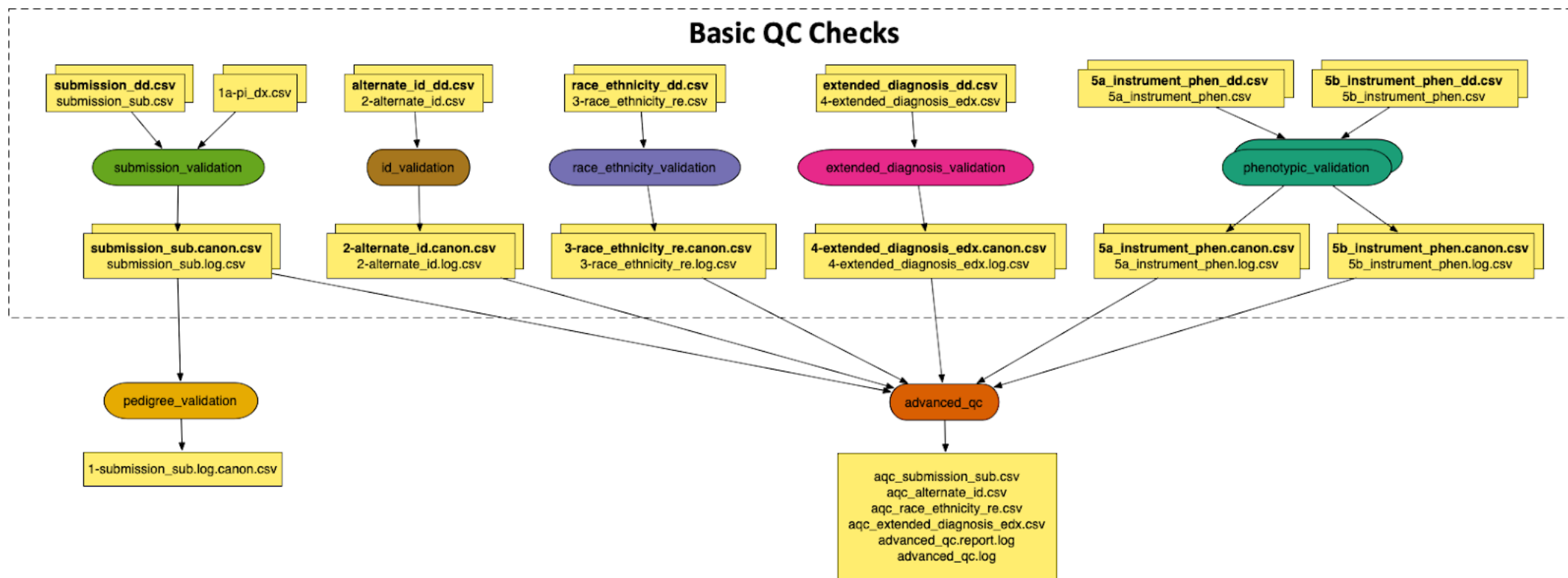
# Introduction

- AutoQC Workflow developed for NIMH Repository & Genomics Resource (NRGR)
- NIMH mandates all data and bio-specimens collected from genomic studies be submitted to a data repository
- Other researchers can get access to these data and bio-specimens
- Example
- AutoQC workflow was developed to check data being submitted by NIMH funded studies

# AutoQC Workflow

- Researchers submit data via a web interface, with real time reporting
- Data is submitted as a collection of csv files
- Each file is associated with a data dictionary describing the data
- The system provides the dictionary for standardized data
- Researcher provides the dictionary for non-standard data
- First – The system checks data against data dictionary
- Last – The system performs advanced checks

# Workflow Structure



# Web Interface

# Previous Validations

[Back to Validate with AutoQC](#)

Showing 1-10 of 3,174 Results

Filter by Keyword

Q

Request ID	Disorder	Email	Submitted?	File	Started On
5e4ec74311098	Obsessive Compulsive Disorder (OCD)	<a href="mailto:gillian.davis@rutgers.edu">gillian.davis@rutgers.edu</a>	Yes	<a href="#">study_138.zip</a>	Feb 20, 2020 09:52 AM
5e334489e4e0f	Autism	<a href="mailto:cbest@lifespan.org">cbest@lifespan.org</a>	Yes	<a href="#">Study200_Files.zip</a>	Jan 30, 2020 01:03 PM
5df2667c06e04	Autism	<a href="mailto:wilson@dls.rutgers.edu">wilson@dls.rutgers.edu</a>	Yes	<a href="#">NJLAGS_final.zip</a>	Dec 12, 2019 08:10 AM
5ddd65eba787b	Bipolar Disorder	<a href="mailto:timothy.barton@nih.gov">timothy.barton@nih.gov</a>	Yes	<a href="#">Study48.zip</a>	Nov 26, 2019 09:50 AM
5dd81a56a6b54	Schizophrenia	<a href="mailto:sallycheung@mail.nih.gov">sallycheung@mail.nih.gov</a>	Yes	<a href="#">Study9_2019.zip</a>	Nov 22, 2019 09:26 AM
5d9b84b157921	Suicide	<a href="mailto:danli.chen@hsc.utah.edu">danli.chen@hsc.utah.edu</a>	Yes	<a href="#">10-1-19 (2).zip</a>	Oct 07, 2019 11:32 AM
5d8a1dc04ae5d	Depression	<a href="mailto:gillian.davis@rutgers.edu">gillian.davis@rutgers.edu</a>	Yes	<a href="#">study_149.zip</a>	Sep 24, 2019 06:44 AM
5d813d1aef2ea	Schizophrenia	<a href="mailto:dlevy@mclean.harvard.edu">dlevy@mclean.harvard.edu</a>	Yes	<a href="#">study_148.zip</a>	Sep 17, 2019 01:07 PM
5d77e0966af9a	Anorexia Nervosa	<a href="mailto:laura_thornton@med.unc.edu">laura_thornton@med.unc.edu</a>	Yes	<a href="#">ANGI_20190910.zip</a>	Sep 10, 2019 10:42 AM
5d517506b0cc2	Schizophrenia	<a href="mailto:gillian.davis@rutgers.edu">gillian.davis@rutgers.edu</a>	Yes	<a href="#">study_107.zip</a>	Aug 12, 2019 07:17 AM
Request ID	Disorder	Email	Submitted?	File	Started On

# Validate with AutoQC

Previous Validations

Help

OVERVIEW

HOW TO VALIDATE AND SUBMIT DATA

SUBMISSION REQUIREMENTS

VALIDATE WITH AUTOQC

Validate your data for sanity checks and quality control.

Choose File

Browse

What data are you submitting?

-- Choose a Disorder --



Study Id

256

Email Notification

email@address.com

Validate



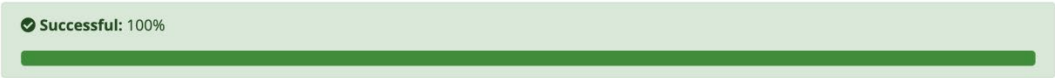


# Auto QC Status

New Validation

Help

[Back to Previous Validations](#)



## Summary

UID	5e6a6ddd95f6e
Disorder	Depression
Study Id	149
File	<a href="#">shaptest7.zip</a>
User	JaclynVitanza
Email	<a href="mailto:jv607@dls.rutgers.edu">jv607@dls.rutgers.edu</a>
Started On	Mar 12, 2020 10:14 AM
Workflow Directory	/web/data/qc/runs/5e6a6ddd95f6e

## Sanity Check Status

Download All Files

File	Submission Validation	Pedigree Validation
study_149_sub.csv	✔ Standardized File	✔ Log

File	ID Validation
study_149_id.csv	✔ Standardized File

File	Phenotypic Validation
shaps01_phen.csv	✔ Standardized File

File	Advanced QC
study_149_sub.canon.csv	✔ Corrected Submission File
study_149_id.canon.csv	✔ Corrected ID File
Corrections Log	✔ Corrections Log
Advanced QC Report	✔ Advanced QC Report



**Thank You!**