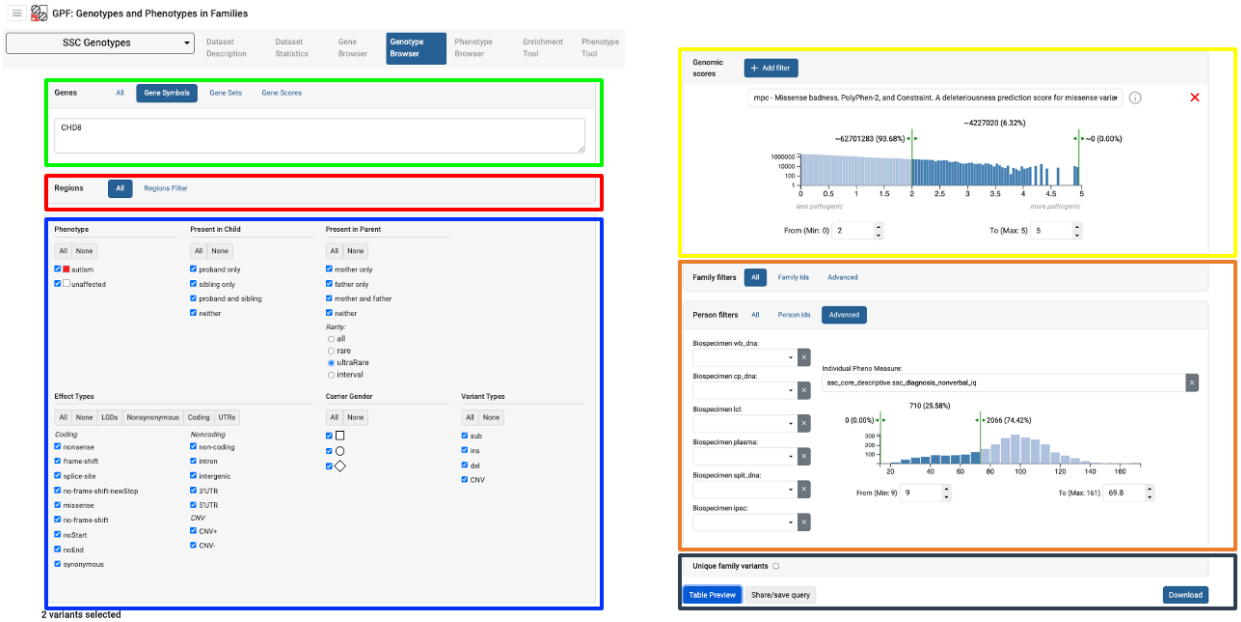
Supplemental Figures

Supplemental Fig 1. Genotype Browser.

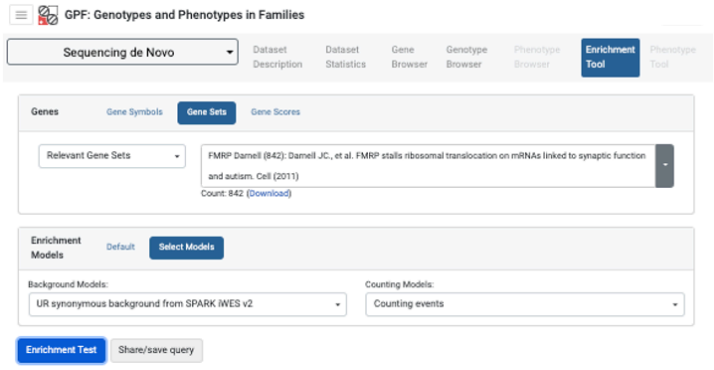


family	variant	genotype	effect	vulnerability/intolerance	phyloP	phastCons	MPC and CADD	FitCons Brain	Frequency	Proband IQs	Races
family id study	location variant	carrier person attributes family person attributes	worst effect genes	LGD rank RVIS rank pLI rank	100way 30way 20way	100way 30way 20way	MPC CADD raw CADD phred	Prefrontal Cortex Male Fetal Female Fetal	SSC exome gnomAD genome gnomAD	NvIQ vIQ	Mom Dad
13838 w1202s766e61	chr14:2140305 sub(A->G)		missense CHD8	83 193 31	9.325 1.307 1.194	1.000 1.000 1.000	3.211 4.522 31.000	0.751 0.751 0.751	- - -	37.0 41.0	white white
13986 EichlerTG2012	chr14:2140997 ins(T)		frame-shift CHD8	83 193 31	5.056 0.681 0.959	1.000 1.000 1.000	2.795 5.470 33.000	0.687 0.687 0.687	- - -	38.0 25.0	white white

The Genotype Browser tab allows users to access the genetic variants associated with a particular dataset. The figure shows one example query and its results in the Genotype Browser tab for the "SSC Genotypes" dataset. The user interacts with the Genotype Browser by first specifying the filters on the properties of the genetic variant of interest and then pressing one of the result buttons (black rectangle), "Table Preview" or "Download," to receive the results into the desired form. The filters of the various types of properties are arranged in separate panels above the result buttons. In the Genes panel (green rectangle), the users can specify the genes or the gene properties (i.e., "Gene Sets" or "Gene Weights") of the genes that are affected by the genetic variants. In the Regions panel (red rectangle), the user can specify the genomic regions of interest. In the next panel (blue rectangle), the user can specify various constraints on the properties of the genetic variant, including the type of variant (i.e., SNVs or CNVs), the effect of the variant on the targeted genes (e.g., missense or synonymous) and the transmission pattern (e.g., present in mother, present in father, *de novo*, present in affected in unaffected children). In the Genomic Scores panel (yellow rectangle), the user can specify constraints on various genomic scores associated with variants (e.g., CADD, polyPhen, MPC). Finally, in the Families and Person filters panels (orange rectangle), the user can specify the families or individuals of interest or constraints on the phenotypic properties (e.g., non-verbal IQ) associated with individuals carrying the genetic variants.

In this example, the user has specified that they are interested only in variants affecting the *CHD8* gene that are ultra-rare (defined as seen only once in the dataset), that have an MPC score larger than 2, and are present in individuals with non-verbal IQ less than 70. The user has pressed the "Tabular Preview" button, resulting in the two genetic variants displayed in a table below the results buttons. Each row of the table describes one variant segregating in one family. The first four columns specify the core properties of the variant: the family where the variant segregates, the definition of the sequence variant (its location and the nucleotide sequence change), the affected gene, and the effect the variant has on the genes. The next columns include properties of the targeted genes (vulnerability scores), properties of the variant itself (the phyloP, phastCons, MPC, CADD, and FitCons brain genomic scores and variant frequencies from different cohorts), and phenotypic properties of the carrier individuals (verbal and non-verbal IQs and father's and mother's races).

Supplemental Fig 2. Enrichment Tool



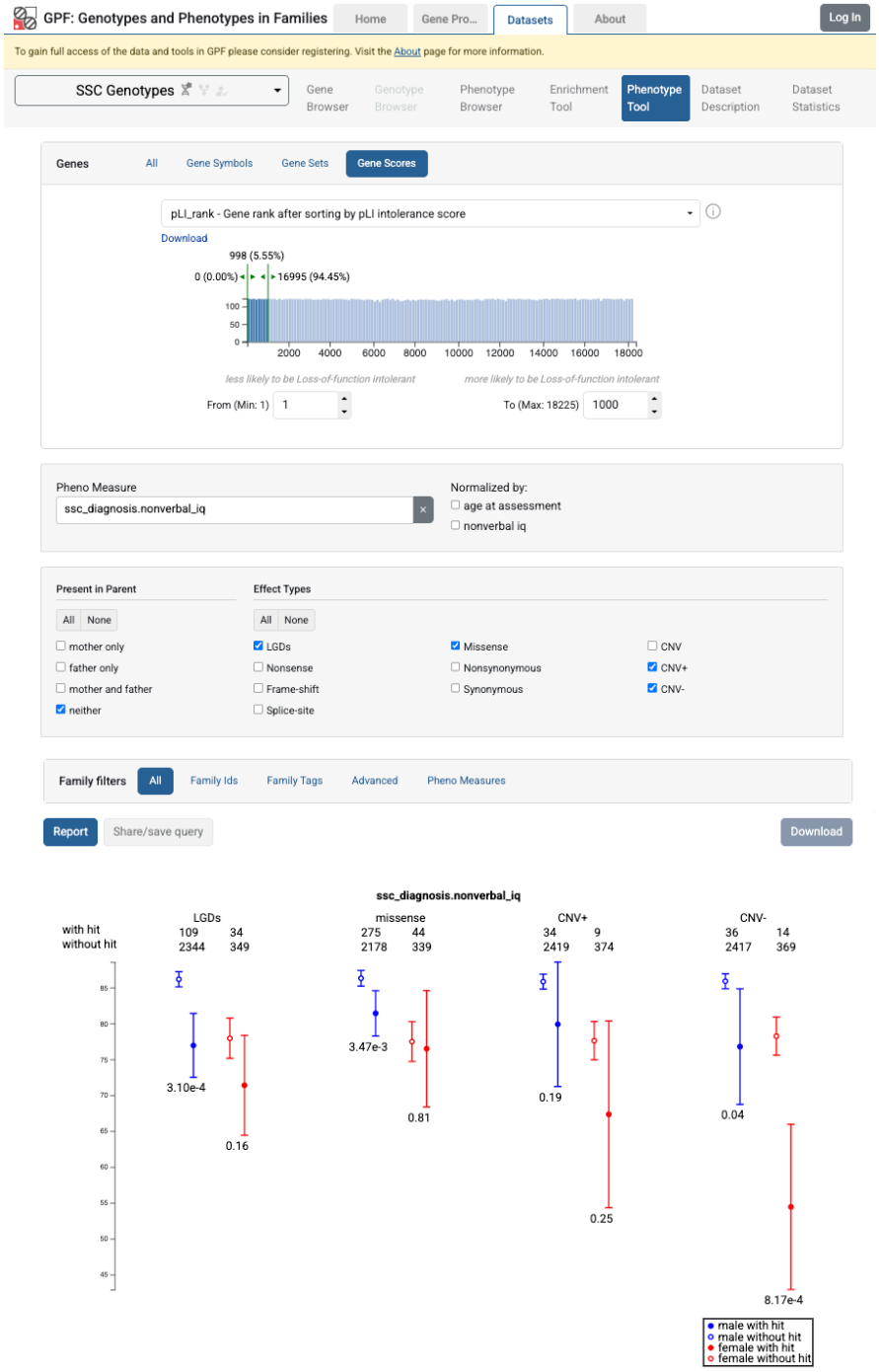
Gene Set: Darnell JC., et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell (2011) (842)

		All				Male				Female			
		N	O	E	pV	N	O	E	pV	N	O	E	pV
autism F: 4327 M: 17206 U: 262	LGDs	3,421	586	347.06	1.36e-35	2,611	419	264.89	1.47e-20	761	155	77.20	5.70e-17
	Missense	20,500	2,387	2,079.73	3.26e-12	16,041	1,813	1,627.36	1.81e-6	4,283	544	434.51	7.86e-8
	Synonymous	8,628	906	875.31	0.2768	6,760	725	685.80	0.1161	1,806	174	183.22	0.5076
congenital heart disease F: 451 M: 762 U: 0	LGDs	140	18	14.20	0.2645	86	11	8.72	0.3739	54	7	5.48	0.4951
	Missense	843	84	85.52	0.9092	512	52	51.94	1.00	331	32	33.58	0.8556
	Synonymous	283	31	28.71	0.6227	171	18	17.35	0.8008	112	13	11.36	0.6371
developmental disorder F: 1654 M: 2010 U: 0	LGDs	1,184	323	120.12	8.47e-62	599	161	60.77	6.24e-31	585	162	59.35	9.56e-33
	Missense	4,668	730	473.57	3.11e-31	2,605	395	264.28	1.42e-15	2,063	335	209.29	1.65e-17
	Synonymous	1,656	180	168.00	0.3286	938	95	95.16	1.00	718	85	72.84	0.1377
epilepsy F: 106 M: 158 U: 254	LGDs	71	20	7.20	1.75e-5	19	4	1.93	0.1197	17	5	1.72	0.0234
	Missense	247	66	25.06	1.53e-13	105	19	10.65	0.0138	93	22	9.43	0.0002
	Synonymous	59	11	5.99	0.0475	34	9	3.45	0.0056	25	2	2.54	1.00
intellectual disability F: 53 M: 47 U: 866	LGDs	243	76	24.65	1.40e-19	3	1	0.30	0.2745	10	3	1.01	0.0727
	Missense	740	129	75.07	1.69e-9	25	6	2.54	0.0355	23	10	2.33	3.78e-5
	Synonymous	240	25	24.35	0.8311	7	0	0.71	1.00	10	1	1.01	1.00
schizophrenia F: 257 M: 742 U: 23	LGDs	103	13	10.45	0.4122	55	5	5.58	1.00	48	8	4.87	0.1465
	Missense	620	94	62.90	0.0001	354	57	35.91	0.0005	266	37	26.99	0.0527
	Synonymous	235	33	23.84	0.0517	149	18	15.12	0.4153	86	15	8.72	0.0316
unaffected F: 3687 M: 3604 U: 14	LGDs	784	74	79.54	0.5542	396	36	40.17	0.5598	385	37	39.06	0.8002
	Missense	6,303	654	639.44	0.5452	3,096	338	314.09	0.1532	3,190	314	323.63	0.5976
	Synonymous	2,728	263	276.76	0.4096	1,401	133	142.13	0.4518	1,320	130	133.91	0.7498

The figure shows an example of using the Enrichment Tool. The Enrichment Tool allows the user to test if a given set of genes is affected by *de novo* mutations more or less than expected in the children in the selected dataset. The user selects the genes to be tested using the Genes panel (green rectangle) and computes expected numbers of *de novo* mutations of three types, LGD, missense, and synonymous, in groups of children from the datasets defined by the primary diagnosis. The Enrichment Tool displays the total number of variants of the particular type for each diagnosis (N columns), the observed number of these variants that affect a gene from the selected genes (O columns), and the expected number of variants in the genes (E). The expected number of variants is computed using a background model that the user can control through the Enrichment Models panel. Finally, the tool computes and displays a p-value (pV columns) for the observed number of variants in the genes based on a Poisson distribution with a parameter equal to the expected number of variants in the selected genes.

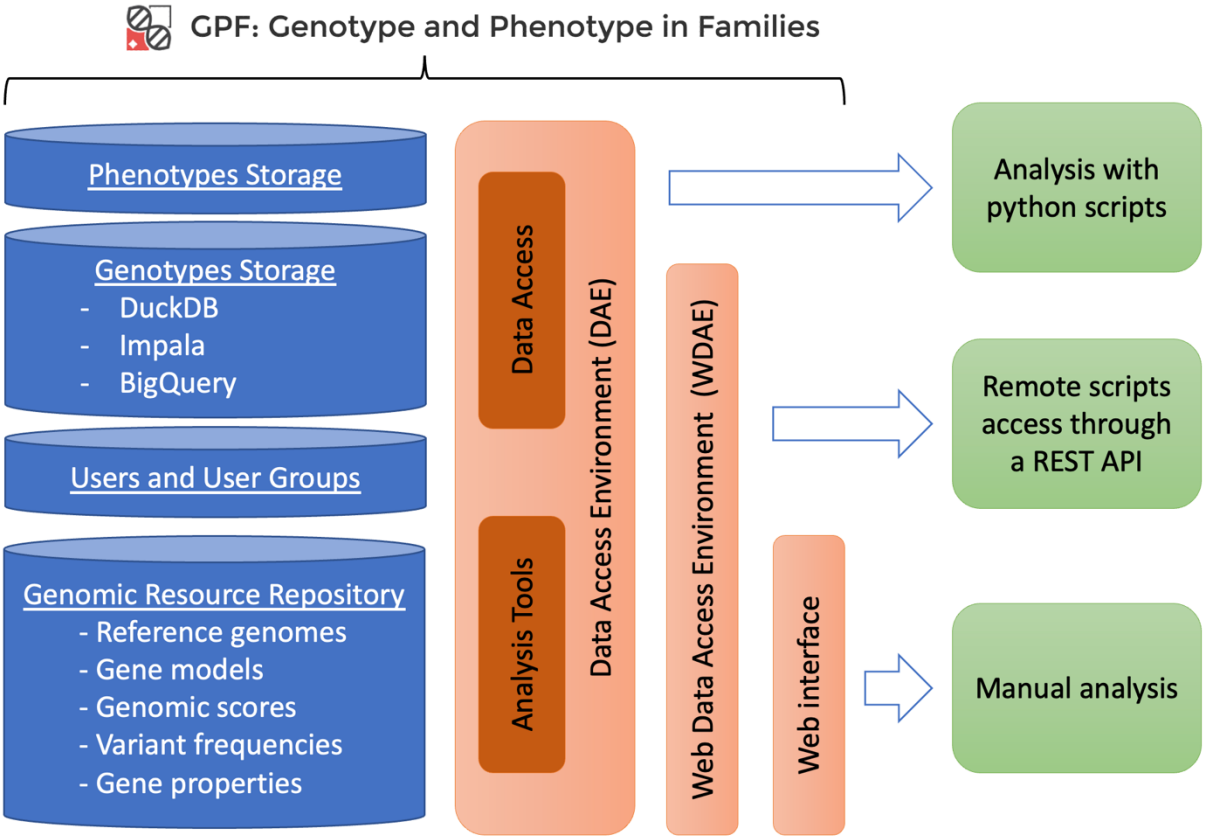
The example above shows the enrichment results in the Sequencing de novo Dataset, including children with six primary diagnoses (autism, congenital heart disease, epilepsy, intellectual disability, schizophrenia, developmental disorder, and unaffected control children) for the set of FMRP target genes. A significant enrichment (red background) is seen for non-synonymous variants in autism, epilepsy, intellectual disability, and schizophrenia but not in congenital heart disease or of the unaffected controls. No synonymous variants in the FMRP target genes are enriched in any phenotype.

Supplemental Fig 3. Phenotype Tool



The Phenotype Tool tests if a given phenotypic measure (i.e., non-verbal IQ) is different between the children that carry the specified type of genetic variant (i.e., *de novo* LGDs) and the children that do not have such variants. The example shows the effect on non-verbal IQ (the nonverbal_iq measure from the ssc_diagnosis instrument, ssc_diagnosis.nonverbal_iq) of the variants in genes with pLI rank of less than or equal to 1000, or the 1000 genes most intolerant to loss-of-function mutations. For example, we see a significant decrease in non-verbal IQ for affected males with de novo LGDs and females with a large de novo deletion (CNV-) affecting one of the selected genes (both with p-value ~ 5×10^{-4}). In addition, large *de novo* duplications (CNV+) exhibit no statistically significant impact in both genders, while the influence of *de novo* missense variants is only marginal in males (p-value = 0.01).

Supplemental Fig 4. GPF Architecture



GPF's architecture consists of three integrated layers. At its core is the Data Access Environment (DAE) layer, which serves as the system's foundation, with access to phenotype, genotype, user, and genomic resource repository information. The second layer introduces the WDAE REST API, enabling remote scripts to interface with DAE's data and analysis tools. The outermost layer is GPF's user web interface, allowing users to execute intricate queries and utilize the power tools outlined in the manuscript.