

## Permutation test scheme

Run the scripts from permutation directory

Check a job in SLURM with `squeue -u <username>`

Step	Sub step	Description	input files	output files (default: in ped_results)	script
1		Generates mutation-sample matrix	gene-sample table, eg <i>projectGeneSample.txt</i>	<i>project_CTs_list.txt</i> <i>matrix_gene_sample.txt</i> (in <i>ped_data</i> )	<i>run_script_data_prep.sh -&gt;</i> <i>data_prep_ped.R</i>
2	non PAN [array]	Generates 1M permuted matrices	<i>project_CTs_list.txt</i> <i>matrix_gene_sample.txt</i>	<i>project/ct/perm_seed_x_ct.rds</i>	<i>batch_run_script.sh -&gt;</i> <i>run_script.sh -&gt;</i> <i>GI_detection_permutation.R</i>
	A (PAN)		<i>project_CTs_list.txt</i> <i>matrix_gene_sample.txt</i>	<i>project/PAN_obj.R</i>	<i>run_script_PAN_A.sh -&gt;</i> <i>GI_detection_permutation_PAN_A.R</i>
	B (PAN) [array]		<i>project/PAN_obj.R</i>	<i>project/PAN/</i> <i>perm_seed_x_PAN.rds</i>	<i>batch_run_script_PAN_B.sh -&gt;</i> <i>run_script_PAN_B.sh -&gt;</i> <i>GI_detection_permutation_PAN_B.R</i>
3	A	Get observed cooccurrence counts for each gene pair	<i>project_CTs_list.txt</i> <i>matrix_gene_sample.txt</i>	<i>project/emp_P_results/ct.R</i>	<i>run_script_emp_P_A.sh -&gt;</i> <i>GI_detection_Empirical_P_A.R</i>
	B [array]	For each RDS file, get cooccurrence counts for each of its permuted matrices and compare with observed counts	<i>project/ct/perm_seed_x_ct.rds</i> <i>project/emp_P_results/ct.R</i>	<i>project/T/ct/x.R</i>	<i>batch_run_script_emp_P_B.sh -&gt;</i> <i>run_script_emp_P_B.sh -&gt;</i> <i>GI_detection_Empirical_P_B.R</i>
	C	Calculate empirical p-value by adding up all comparisons from step B	<i>project/emp_P_results/ct.R</i> <i>project/T/ct/x.R</i>	<i>project/emp_P_results/ct.txt</i>	<i>run_script_emp_P_C.sh -&gt;</i> <i>GI_detection_Empirical_P_C.R</i>
4	A	Randomly pick a matrix, repeat 100 times	<i>project/ct/perm_seed_x_ct.rds</i>	<i>project/P/ct/perm_mtx.R</i>	<i>run_script_null_dist_A.sh -&gt;</i> <i>null_dist_A.R</i>
	B [array]	Compare sampled mtxs with mtxs in RDS file	<i>project/ct/perm_seed_x_ct.rds</i> <i>project/P/ct/perm_mtx.R</i>	<i>project/P/ct/perm_co_mtx_x.R</i>	<i>batch_run_script_null_dist_B.sh -&gt;</i> <i>run_script_null_dist_B.sh -&gt;</i> <i>null_dist_B.R</i>
	C	Merge all data and write p-values to file	<i>project/P/ct/perm_co_mtx_x.R</i>	<i>project/emp_P_results/null_dist/</i> <i>ct/come_seed_x.txt</i>	<i>run_script_null_dist_C.sh -&gt;</i> <i>null_dist_C.R</i>
5		Calculate the empirical FDR given the observed p-values and a null distribution of p-values	<i>project/emp_P_results/ct.txt</i> <i>project/emp_p_results/null_dist/ct/</i> <i>come_seed_x.txt</i>	<i>project/emp_P_results/ct/</i> <i>ct.txt_emp_FDR.txt</i>	<i>run_script_FDR_correction.sh -&gt;</i> <i>FDR_correction.R</i>