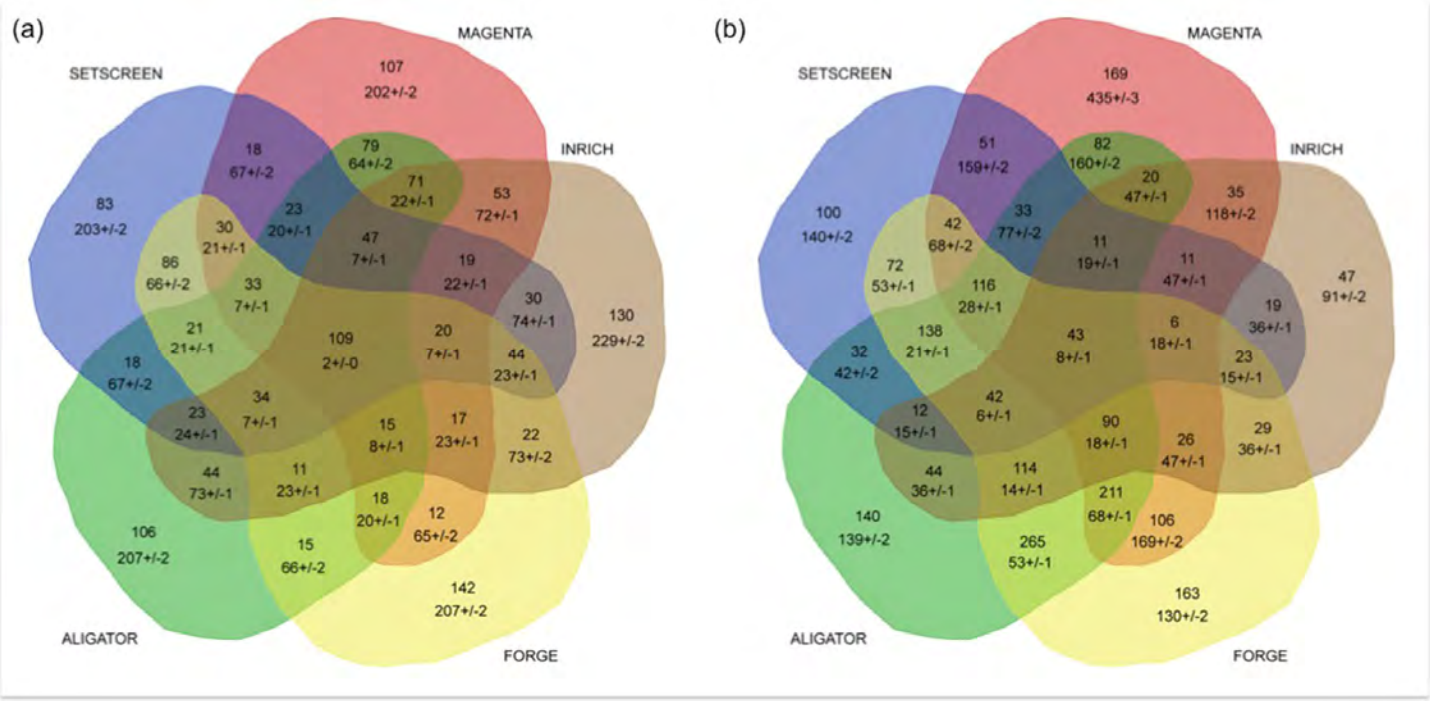
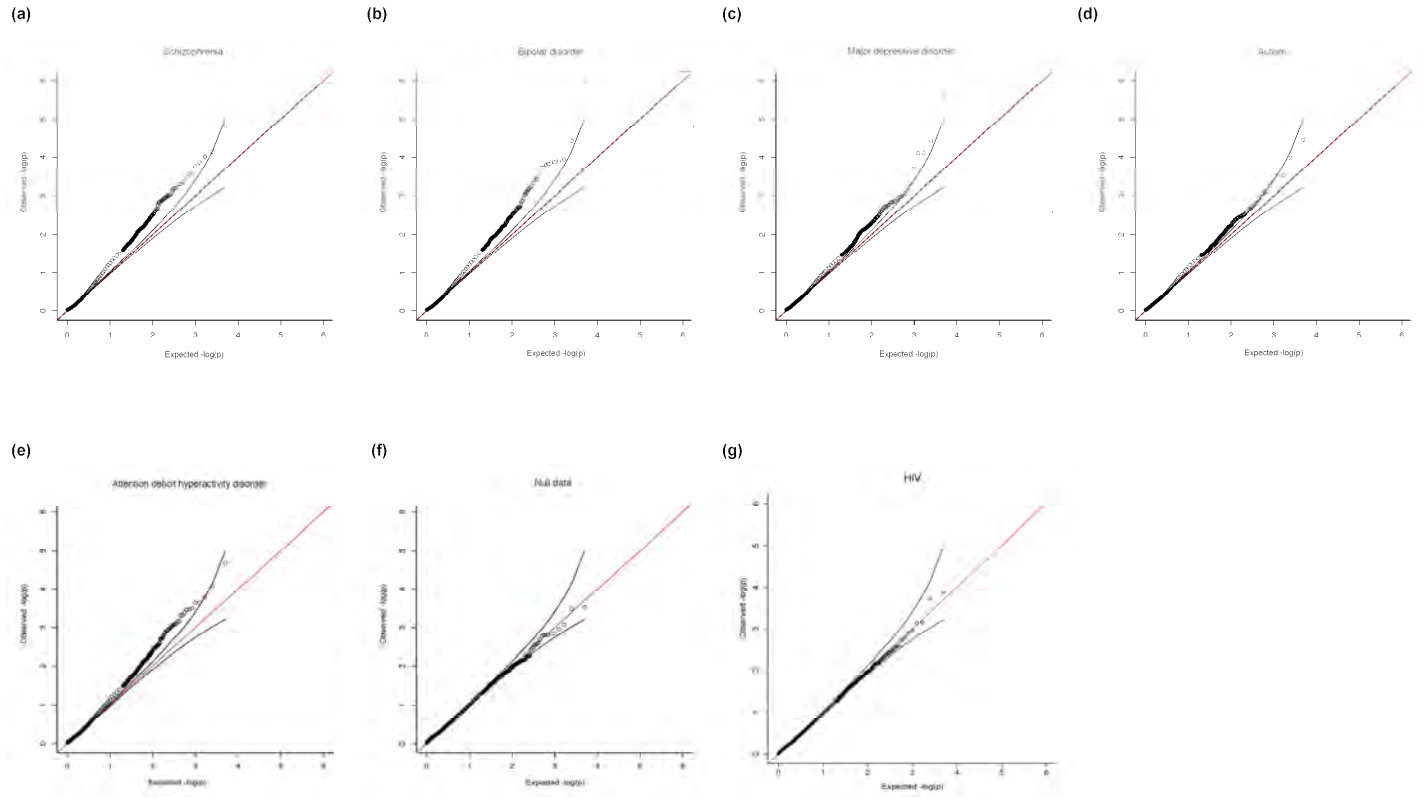


Supplementary Figure 1

Observed and expected (shown as ranges representing 95% confidence intervals) number of significant pathways overlapping between different sets of results from different methods.

(a) Analysis conducted using the SCZ dataset, top 25% of each method. (b) Analysis conducted using the NULL dataset, top 25% of each method.



Supplementary Figure 2

Quantile-quantile plots summarizing combined p-values for each method on different phenotypes.

Order is (a) SCZ, (b) BIP, (c) MDD, (d) AUT, (e) ADD, (f) null data from permutations and, (g) HIV.

Psychiatric genomewide association study analyses implicate neuronal, immune and histone pathways.

Table of Contents

Supplemental Analysis

Gene set summaries

Pathway/ gene set	# sets	Preparation	URL
KEGG	237	Uniprot file was reduced to remove redundant IDs and converted to Entrez, Ensembl and HGNC IDs; Ensembl file was not edited after download	ftp://ftp.genome.jp/pub/kegg/genes/organisms/hsa/hsa_ensembl-hsa.list ftp://ftp.genome.jp/pub/kegg/genes/organisms/hsa/hsa_pathway.list ftp://ftp.genome.jp/pub/kegg/genes/organisms/hsa/hsa_uniprot.list
GO	14793	Uniprot files were saved and Uniprot IDs were converted to Ensembl, Entrez and HGNC IDs. A script was run to recursively grab parent terms for all terms defined in the downloaded file. The result has the following format: Gene|Term|X, Where 'X' is either direct or indirect, where indirect represents the parent of a parent	http://cvsweb.geneontology.org/cgiAbin/cvsweb.cgi/go/geneAassociations/gene_association.goa_human.gz?rev=HEAD http://www.geneontology.org/ontology/obo_format_1_2/gene_ontology_ext.obo (parent tems - "is_a", "part_of")
PANTHER	140	"HUMAN" was parsed and file was saved in Ensembl and Uniprot ID. Uniprot was converted to Entrez and HGNC IDs	ftp://ftp.pantherdb.org/pathway/current_release/SequenceAssociationPathway3.01.txt
TARGETSCAN	162	Files were downloaded in HGNC ID format and parsed for "human". Pathways were determined by a context score cutoff of < -0.3 and converted to Uniprot, Ensembl and Entrez IDs	http://www.targetscan.org//vert_50//vert_50_data_download/Summary_Counts.txt.zip
OMIM	4380	HGNC file was saved and converted into Ensembl IDs	ftp://ftp.ncbi.nih.gov/repository/OMIM/ARCHIVE/morbidmap
REACTOME	40	Files were grepped for human and Uniprot IDs were converted into Entrez, Ensembl and HGNC IDs	http://www.reactome.org/download/current/curated_and_inferred_uniprot_2_pathways.txt

Table S1. Summary of preparation of various pathway/gene sets for analysis. Pathways considered contained 2-200 genes. Ensembl ids were used as the master gene set, with -35kb upstream and +10kb downstream coordinates used to define the gene boundaries (HG 19; NCBI build 36.1). When multiple transcripts mapped to a single gene locus, the longest transcript was used as a representative.

Table S2. Summary statistics for all pathway/gene sets used.

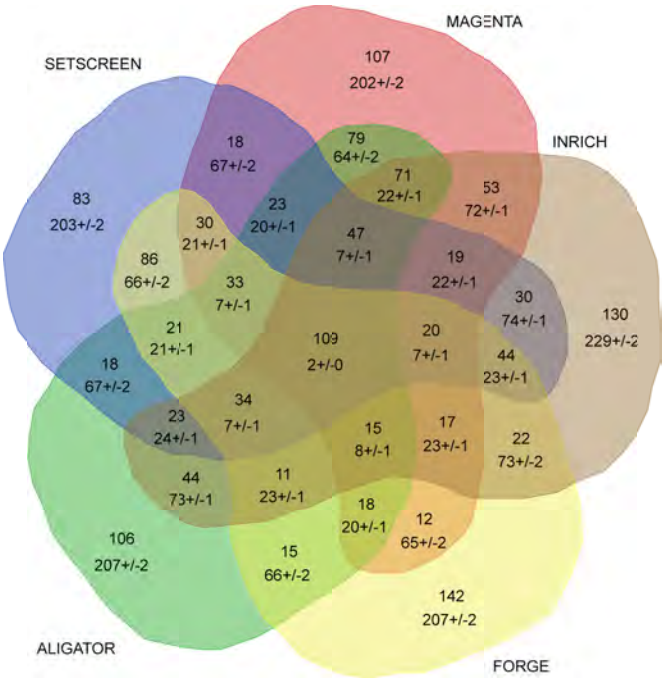
	Num. of Pathways Downloaded	Num. of Genes	Num. of Pathways in Restricted Set	Num. of Genes Covered in	Average Pathway Length	Median Pathway Length

		Covered	[10-200]	Restricted Set [10-200]	(restricted set)	(restricted set)
KEGG	237	5946	214	5202	65.49	54
GO	14793	17764	4550	13794	44.93	26
PANTHER	140	2207	72	1628	46.12	37.5
TARGETSCAN	162	10901	77	4563	84.64	76
REACTOME	40	1518	28	906	59.82	52
OMIM	4380	3170	8	97	12.5	12.5
Total (unique)	19752	18848	4949	14593		

Comparisons between methods: method overlap

The results obtained from this procedure are represented in Figure S2 as a 5-way Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/). Each overlap quantity is associated with an expected value that is derived by random sampling with no correlation among methods. We conclude that top pathways from the five methods overlap more than expected by chance. For example, in the SCZ dataset, there are 109 independent pathways shared across the top 25% of all methods, while only 2 are expected by chance (Figure S2(a)). Similar patterns were seen for the NULL dataset (Figure S2(b)).

(a)



(b)

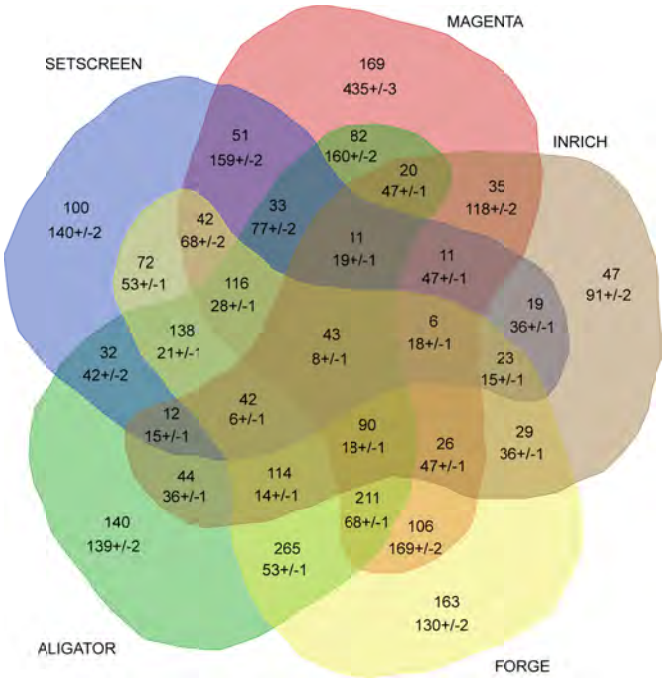


Figure S1. **Observed and expected (shown as ranges representing 95% confidence intervals) number of significant pathways overlapping between different sets of results from different methods.** (a) Analysis conducted using the SCZ dataset, top 25% of each method. (b) Analysis conducted using the NULL dataset, top 25% of each method.

P-value distributions for all disorders

We ran 5 methods on each of the 5 phenotypes and on a null dataset. The method-wise p-values were then combined as described above. QQ plots summarize the p-value distribution.

HERE

Figure S2. **Quantile-quantile plots summarizing combined p-values for each method on different phenotypes.** Order is (a) SCZ, (b) BIP, (c) MDD, (d) AUT, (e) ADD, (f) null data from permutations and, (g) HIV.

Table S3. **Pearson correlations of pathway enrichments between pairs of methods**. Calculated on 1,918 pathways with Jaccard coefficient <0.2, averaged across five psychiatric disorders (averaging performed by calculating the mean of the z-transformed correlation coefficients for the individual disorders, weighted by the inverse of the variance). All correlations nominally significant except Forge-Magenta and Magenta-Aligator on the null dataset.

	Forge	**Magenta**	**Inrich**	**Aligator**
Setscreen	0.653, 9543	0.321, 9340	0.278, 2528	0.355, 3824
Forge		0.213, 9293	0.137, 2526	0.187, 3823
Magenta			0.335, 2442	0.468, 3705
Inrich				0.566, 2288

Ranked pathways across individual disease groups

Combining pathway enrichments across diseases

	AUT	BIP	HIV	MDD	NULL	SCZ
ADHD	0.170	0.234	0.126	0.192	0.119	0.182
AUT		0.238	0.125	0.176	0.114	0.246
BIP			0.102	0.201	0.127	0.290
HIV				0.181	0.166	0.092
MDD					0.111	0.254
NULL						0.156

Table S4. Pearson correlation of pathway-specific enrichment p-values between pairs of datasets. Calculated on 1,918 Jaccard-pruned pathways.

Table S5. Pathway results for each disease/cross-disease dataset.

See excel file "Table_S5.xls".

	AUT	BIP	HIV	MDD	NULL	SCZ
ADHD	P=0.044	P<0.001	P=0.619	P=0.002	P=0.732	P=0.019
AUT		P<0.001	P=0.625	P=0.029	P=0.789	P<0.001
BIP	P=0.244		P=0.912	P<0.001	P=0.636	P<0.001
HIV	P=0.993	P=0.855		P=0.016	P=0.080	P=0.969
MDD	P=0.989	P=0.166	P=0.749		P=0.837	P<0.001
NULL	P=0.215	P=0.862	P=0.453	P=0.933		P=0.158
SCZ	P=0.015	P=0.003	P=0.787	P=0.006	P=0.788	

Table S6. Significance of (a) Pearson correlation of pathway-specific enrichment p-values (top triangle) and (b) overlap in top 10% of enriched pathways (bottom triangle, shaded) between datasets, allowing for the mean correlation in pathway-specific enrichment p-values observed with the null dataset (0.132). Calculated on 1,918 Jaccard-pruned (<=0.2) pathways.

Ranked pathways across multiple individual disease groups

Table S7. Ranked list of pathways when combined across 3 well-powered disease groups.

See excel file "Table_S7.xls".

Table S8. Ranked list of pathways when combined across SCZ, HIV and NULL datasets.

See excel file "Table_S8.xls".

Table S9. Details related to module-level summaries of spatial, temporal, and cell-type specific expression patterns.

See excel file "Table_S9.xls".

Table S10. Ranked list of pathways when combined across all 5 disease groups.

See excel file "Table_S10.xls".

Table S11. P-values for the SNPs in each gene in each disorder and their pathways membership are presented for all pathways with q<0.1.

See excel file "Table_S11.xls".

Ranking pathways using the four methods giving the highest ranks

Although pathway enrichment p-values from the five analysis methods are significantly correlated with each other (see Table S3), it is possible that one of the methods may differ from the others for any given pathway. Such a discrepancy may reduce the average rank of the pathway and thus its significance. For example, a pathway with ranks (1, 1, 1, 1, 996) has the same average rank as one with ranks (200, 200, 200, 200, 200), but should be considered as more important. To increase power to detect enrichment to such pathways, we performed pathway enrichment analyses as described previously using the average of the four highest ranks for each pathway (discarding the lowest). A comparison of the enrichment p-values obtained by this approach to those based on all five methods are given in Table S11 for the significant pathways for individual diseases reported in Table 1, and in Table S12 for the 15 pathways with q<0.05 in the combined SCZ+BIP+MDD analysis (Table 2). Using the four highest ranks in general had little effect on the enrichment p-values for these

pathways, although the three most significant pathways in schizophrenia, all related to the synapse, had q-values <0.05. Furthermore, using the four highest ranks discovered no other significant pathways.

P rank (all 5)	q-value (all 5)	p rank (best 4)	q-value (best 4)	Pathway ID	Description
BIP					
1.01E-06	0.005	9.70E-06	0.024	GO:51568	histone H3-K4 methylation
3.82E-05	0.093	4.65E-06	0.023	path:hsa05218	Melanoma
1.16E-04	0.093	5.06E-04	0.209	GO:7129	(chromosomal) synapsis
1.27E-04	0.093	3.06E-04	0.191	path:hsa05213	Endometrial cancer
1.34E-04	0.093	2.54E-04	0.191	P00003	Alzheimer_disease-amyloid_secretase_pathway
1.35E-04	0.093	1.73E-04	0.191	path:hsa05215	Prostate cancer
1.50E-04	0.093	4.19E-04	0.209	path:hsa05216	Thyroid cancer
1.59E-04	0.093	8.93E-04	0.218	GO:90066	regulation of anatomical structure size
1.81E-04	0.093	1.97E-04	0.191	path:hsa05214	Glioma
1.87E-04	0.093	7.96E-04	0.218	GO:70192	chromosome organization involved in meiosis
SCZ					
1.58E-05	0.078	2.73E-05	0.045	GO:14069	postsynaptic density
7.15E-05	0.160	2.32E-05	0.045	GO:45211	postsynaptic membrane
9.67E-05	0.160	2.48E-05	0.045	GO:43197	dendritic spine
1.36E-04	0.168	1.94E-04	0.192	GO:51568	histone H3-K4 methylation
1.74E-04	0.173	4.85E-04	0.291	GO:33267	axon part
MDD					
2.63E-06	0.012	2.28E-05	0.111	GO:8601	protein phosphatase type 2A regulator activity
3.88E-05	0.092	2.38E-04	0.165	GO:34330	cell junction organization
7.70E-05	0.094	1.82E-04	0.165	GO:43297	apical junction assembly
7.92E-05	0.094	8.14E-05	0.165	GO:45216	cell-cell junction organization
1.97E-04	0.186	2.38E-04	0.165	GO:31056	regulation of histone modification

Table S12. Effect of enrichment analysis based on the average of the four highest ranks on the top pathways from BIP, SCZ and MDD (Table 1).

Rank	Combined p (all 5)	q-value (all 5)	Combined p (best 4)	q-value (best 4)	Pathway ID	Description
1	5.75E-08	0.0003	2.67E-07	0.0013	GO:51568	histone H3-K4 methylation
2	1.46E-05	0.0362	3.08E-05	0.0336	GO:16571	histone methylation
3	4.73E-05	0.0414	9.36E-05	0.0515	GO:43414	macromolecule methylation
4	5.10E-05	0.0414	8.21E-05	0.0508	GO:34968	histone lysine methylation
5	5.58E-05	0.0414	1.48E-04	0.0709	GO:45216	cell-cell junction organization
6	5.69E-05	0.0414	1.89E-05	0.0336	P00003	Alzheimer_disease-amyloid_secretase_pathway
7	5.86E-05	0.0414	6.72E-05	0.0475	P04393	Ras_Pathway
8	7.12E-05	0.0422	2.73E-04	0.0712	GO:8601	protein phosphatase type 2A regulator activity
9	7.83E-05	0.0422	1.97E-04	0.0712	GO:43297	apical junction assembly
10	9.25E-05	0.0422	6.08E-05	0.0475	P00052	TGF-beta_signaling_pathway
11	9.53E-05	0.0422	2.85E-04	0.0712	GO:14069	postsynaptic density
12	0.0001	0.0422	1.58E-04	0.0709	GO:32869	cellular response to insulin stimulus
13	0.0001	0.0450	3.43E-04	0.0712	P00010	B_cell_activation
14	0.0001	0.0450	2.17E-04	0.0712	GO:8757	S-adenosylmethionine-dependent methyltransferase activity
15	0.0001	0.0454	2.06E-03	0.1725	GO:23061	signal release
16	0.0002	0.0473	9.67E-04	0.1328	GO:34330	cell junction organization

Table S13. Effect of enrichment analysis based on the average of the four highest ranks on the 16 pathways with q<0.05 from a combined analysis of BIP, SCZ and MDD (Table 2).

Robustness of enrichment analyses to variation in correlation between methods.

To obtain an accurate significance level for the average rank across methods for each pathway, it is necessary to model the dependence between the analysis methods. We did this by calculating the correlation in p-values between each pair of methods based on the null dataset (such that correlations measure only the dependence between methods and not the presence of true signals that would induce correlation between methods). These correlations are given in Table S13, together with the numbers of independent pathways with Jaccard coefficient <0.2 from which they were derived. This set of correlation coefficients were then used to generate sets of correlated uniform variants to represent enrichment p-values, which were then turned into simulated sets of ranks to obtain empirical p-values for the average ranks for each pathway in the observed data. The correlations given in Table S13 are only estimates of the true correlation, and as such susceptible to stochastic variation. To investigate the effect of variation in correlation between methods on the pathway enrichments, we calculated 95% confidence intervals for each of the correlations given in Table S13 (using the number of

pathways from which they were generated). Then, we performed two enrichment analyses using the same methodology as describe earlier: one with each of the correlation coefficients between methods set to the lower bound of the 95% confidence interval (or zero, if this was negative) and one with the correlation coefficients set to the upper bound of the 95% confidence interval. In the first analysis, dependence between methods is likely to be underestimated, resulting in increased significance for pathways with high average rank across methods. Conversely, in the second analysis, dependence between methods is overestimated, thus reducing significance of pathways with high average rank, and thus giving a conservative analysis. The first analysis was applied to the null dataset and the combined analysis of the null, schizophrenia and HIV datasets, in order to check that no false-positive results were generated (after correcting for multiple testing of pathways). The second analysis was performed on the individual disease samples, as well as the analyses combining schizophrenia, bipolar and major depression, in order to test whether the results are still significant when a conservative assumption is made about the dependence between analysis methods. Results are given in Table S14 for the significant pathways for individual diseases reported in Table 1, and in Table S15 for the 15 pathways with q<0.05 in the combined SCZ+BIP+MDD analysis (Table 2).

As Tables S14 and S15 show, making conservative assumptions about the dependency between analysis methods has little effect on the significance of the top pathways. Likewise, underestimating the dependency between methods does not induce false positives in the null dataset (min q-value=0.187) or the combined SCZ+HIV+null analysis (min q=0.075). Thus, the analyses presented in this paper are robust to assumptions regarding the extent of correlation between the analysis methods.

	Forge	**Magenta**	**Inrich**	**Aligator**
Setscreen	0.659, 1798	0.068, 1868	0.205, 959	0.407, 802
Forge		0.025, 1749	0.141, 874	0.328, 721
Magenta			0.290, 933	0.091, 778
Inrich				0.351, 721

Table S14. Pearson correlations of pathway enrichments between pairs of methods on the null dataset, calculated on 1,918 pathways with Jaccard coefficient <0.2, The number of pathways used to calculate each correlation is also given.

P rank (original)	**q-value (original)**	**p rank (upper 95)**	**q-value (upper 95)**	**Pathway ID**	**Description**

BIP					
1.01E-06	0.005	2.22E-06	0.011	GO:51568	histone H3-K4 methylation
3.82E-05	0.093	4.28E-05	0.086	path:hsa05218	Melanoma
1.16E-04	0.093	1.15E-04	0.086	GO:7129	(chromosomal) synapsis
1.27E-04	0.093	1.28E-04	0.086	path:hsa05213	Endometrial cancer
1.34E-04	0.093	1.28E-04	0.086	P00003	Alzheimer_disease-amyloid_secretase_pathway
1.35E-04	0.093	1.30E-04	0.086	path:hsa05215	Prostate cancer
1.50E-04	0.093	1.43E-04	0.086	path:hsa05216	Thyroid cancer
1.59E-04	0.093	1.50E-04	0.086	GO:90066	regulation of anatomical structure size
1.81E-04	0.093	1.76E-04	0.086	path:hsa05214	Glioma
1.87E-04	0.093	1.81E-04	0.086	GO:70192	chromosome organization involved in meiosis
SCZ					
1.58E-05	0.078	2.16E-05	0.107	GO:14069	postsynaptic density
7.15E-05	0.160	8.66E-05	0.172	GO:45211	postsynaptic membrane
9.67E-05	0.160	1.09E-04	0.172	GO:43197	dendritic spine
1.36E-04	0.168	1.39E-04	0.172	GO:51568	histone H3-K4 methylation
1.74E-04	0.173	1.80E-04	0.178	GO:33267	axon part
MDD					
2.63E-06	0.012	6.46E-06	0.029	GO:8601	protein phosphatase type 2A regulator activity
3.88E-05	0.092	4.67E-05	0.094	GO:34330	cell junction organization
7.70E-05	0.094	8.08E-05	0.094	GO:43297	apical junction assembly
7.92E-05	0.094	8.28E-05	0.094	GO:45216	cell-cell junction organization
1.97E-04	0.186	1.90E-04	0.172	GO:31056	regulation of histone modification

Table S15. Effect of setting correlations between methods to the upper bound of the 95% confidence interval (thereby overestimating dependence between methods) on the enrichment analysis of the top pathways from BIP, SCZ and MDD (Table 1).

Rank	Combined p (original)	q-value (original)	Combined p (upper 95)	q-value (upper 95)	Pathway ID	Description
1	5.75E-08	0.0003	1.13E-07	0.0006	GO:51568	histone H3-K4 methylation

2	1.46E-05	0.0362	1.53E-05	0.0378	GO:16571	histone methylation
3	4.73E-05	0.0414	5.18E-05	0.0485	GO:43414	macromolecule methylation
4	5.10E-05	0.0414	5.99E-05	0.0485	GO:34968	histone lysine methylation
5	5.58E-05	0.0414	6.26E-05	0.0485	GO:45216	cell-cell junction organization
6	5.69E-05	0.0414	6.38E-05	0.0485	P00003	Alzheimer_disease-amyloid_secretase_pathway
7	5.86E-05	0.0414	6.85E-05	0.0485	P04393	Ras_Pathway
8	7.12E-05	0.0422	1.45E-04	0.0553	GO:8601	protein phosphatase type 2A regulator activity
9	7.83E-05	0.0422	8.89E-05	0.0550	GO:43297	apical junction assembly
10	9.25E-05	0.0422	1.16E-04	0.0553	P00052	TGF-beta_signaling_pathway
11	9.53E-05	0.0422	1.28E-04	0.0553	GO:14069	postsynaptic density
12	0.0001	0.0422	1.25E-04	0.0553	GO:32869	cellular response to insulin stimulus
13	0.0001	0.0450	1.56E-04	0.0553	P00010	B_cell_activation
14	0.0001	0.0450	1.38E-04	0.0553	GO:8757	S-adenosylmethionine-dependent methyltransferase activity
15	0.0001	0.0454	1.82E-04	0.0582	GO:23061	signal release
16	0.0002	0.0473	1.88E-04	0.0582	GO:34330	cell junction organization

Table S16. Effect of setting correlations between methods to the upper bound of the 95% confidence interval (thereby overestimating dependence between methods) on the enrichment analysis of the 16 pathways with $q<0.05$ from a combined analysis of BIP, SCZ and MDD (Table 2).

Table S17. Information related to the supervised network analysis of 797 genes, specifying gene set membership from the 3 major pathways, module membership, and relative connectivity (kME) to each module.

See excel file "Table_S17.xls".