



OPEN

SUBJECT AREAS:

DATA MINING
FUNCTIONAL CLUSTERINGReceived
2 September 2013Accepted
6 February 2014Published
25 February 2014Correspondence and
requests for materials
should be addressed to
H.Q. (quh@mail.cbi.
pku.edu.cn)* These authors
contributed equally to
this work.

A systems biology approach to identify intelligence quotient score-related genomic regions, and pathways relevant to potential therapeutic treatments

Min Zhao*, Lei Kong* & Hong Qu

Center for Bioinformatics, State Key Laboratory of Protein and Plant Gene Research, College of Life Sciences, Peking University, Beijing 100871, P.R. China.

Although the intelligence quotient (IQ) is the most popular intelligence test in the world, little is known about the underlying biological mechanisms that lead to the differences in human. To improve our understanding of cognitive processes and identify potential biomarkers, we conducted a comprehensive investigation of 158 IQ-related genes selected from the literature. A genomic distribution analysis demonstrated that IQ-related genes were enriched in seven regions of chromosome 7 and the X chromosome. In addition, these genes were enriched in target lists of seven transcription factors and sixteen microRNAs. Using a network-based approach, we further reconstructed an IQ-related pathway from known human pathway interaction data. Based on this reconstructed pathway, we incorporated enriched drugs and described the importance of dopamine and norepinephrine systems in IQ-related biological process. These findings not only reveal several testable genes and processes related to IQ scores, but also have potential therapeutic implications for IQ-related mental disorders.

Each human individual differs from others in his complex intelligent capability to think and reason abstractly, remember and learn from experience, solve problems effectively to adapt to the environment, and communicate by exchanging thought¹. It is reported that the interaction of genetic and environmental factors are significantly correlated with human intelligence. A study of twins with highly educated parents showed that genetics accounts for 56% of the variability in reading scores and the shared environment accounts for 12%². While this study is evidence for highly heritable human intelligence, it is single gene-centric, and these studies are typically not consistent across different occasions, different domains, and different criteria. Moreover, many studies of candidate genes for human intelligence cannot be successfully replicated between studies and populations³. In addition, single gene-centric studies cannot present the complexity of cellular interactions and pathways regulating neurons in cognitive processes, which often change according to various spatial, temporal, and environmental factors in the cell. Therefore, single gene-centric studies might also ignore the blueprint of molecules involved in complex cognitive activities, for instance, those genes that express differently due to a single mutation may not be detected, or those genes with lethal effects may be overlooked before those effects appear. The Intelligence Quotient (IQ) test is one of the most widely used phenotypes to measure human intelligence⁴. As a quantitative trait, the heritability associated with a single gene responsible for a small effect on observed IQ score appears mixed^{3,5,6}. Systematic collection and analysis of genetic predisposition information with single gene-centric evidence might provide a clearer biological picture for IQ and intelligence. Our previous literature-based collection and expansion of genes involved in IQ scores present an overview for the genetic basis associated with IQ⁷. Pathway analysis reveals that the gene content associated with IQ score is often related to numerous diseases, such as schizophrenia, autism, depression, and breast cancer⁷⁻¹⁰. The goal of network biology is to clarify the mechanisms ruling the cellular systems, and to create testable hypotheses, which can be used to predict the possible outputs of various perturbations and help to identify therapeutic and diagnostic molecules¹¹. To address high genetic heterogeneities of intelligence and understand biological mechanisms for cognition regardless of environmental factors, we adopted a systems biology-based approach to connect the IQ-related genes with literature evidence. We not only found the genomic distribution, but also reconstructed a connected pathway map for IQ-related genes. Based on the hub nodes in the constructed map, we expanded the protein-



coding based network to a protein-drug interaction network to generate potential and highly enriched drugs for the genes in our reconstructed IQ pathway.

Results

IQ-related genes are enriched in multiple regions of chromosomes 7 and X. To address the genetic complexity of IQ score, we developed IQdb (<http://IQdb.cbi.pku.edu.cn>), a publicly available database for exploring IQ-associated human genes⁷. In IQdb, 158 experimentally verified genes from the literature serve as the core dataset, among which approximately 16% of the IQ-related genes are located on X chromosome. It is reported that only 3.4% of all human genes belong to X chromosome¹². The obvious contrast may indicate that the IQ-related genes are over-represented on X chromosome. To clarify in more detailed which genomic regions contain greater numbers of IQ-related genes, we adopted a hypergeometric test based on the gene content of each cytoband on the human genome. The further enrichment analysis identified 10 genomic regions with significantly larger amount of IQ-related genes. Among 10 regions, 4 belong to X chromosome, including Xp, Xq, Xq28, and Xp11. A previous gene expression study showed that the genes on X chromosome are more likely to express in brain regions and reproductive tissues¹³. Our results may be consistent with the idea that genes on X chromosome may be associated with the quantitative IQ score¹². Since X chromosome is important to sexual differentiation, the enrichment of IQ-related genes on X chromosome may provide an incentive to explore differences in the IQ scores between the sexes^{14,15}.

Previous studies have identified 46 genomic regions involved in IQ using a linkage approach⁷. Though none of these regions are located on X chromosome, six of them are from chromosome 7 in humans (13.04%). Among the ten genomic regions with enriched IQ-related genes, six are not on X chromosome, including 7q, 7q11, 7q31, 8p12, 15q14, and 17p13. Except for 8p12¹⁶, none of these genomic regions overlap with the genomic regions in the prior linkage studies. Three of six regions are from chromosome 7, accounting for 8.86% of the 158 IQ-related genes (Table 1). One previous study showed that 7q31–36 may be linked to verbal IQ based on 361 Australian and Dutch twins¹⁷. Our results may uncover complementary roles of the genes on chromosome 7 in general intelligence.

In summary, our analysis narrows the probable genomic regions to several likely candidates, providing a better understanding of an IQ-related genomics and a highly rational way to interpret differences such as gender. Notably, our gene enrichment based approach has identified multiple genomic regions in chromosome X, which was missed in previous linkage studies. This difference may be caused by the sample size of the population in linkage studies. The other potential reason is that most of genes on chromosome X in IQdb are collected based on single-gene-based functional studies. Genetic association is emerging as a large scale screening tool. These small proportions of genomic regions have permitted global examination of the genome/proteome on a larger population.

The enriched target genes in 158 IQ-related genes for transcription factors and miRNAs related to mental disorders. A fundamental problem in biological systems is discovering potential regulators for candidate genes, which may help us understand the entire volume of genetic information. Compared to enriched IQ-related genes, regulation does not teach us about the linear relationship with the physical chromosome, but the dynamic mechanisms of underlying environmental change.

The modifying effects of several types of regulatory genes are widely studied and can be predicted based on sequence features of their potential target genes. Transcription factor (TF) can influence gene expression through transcription activation or suppression of target genes with different binding efficiency in promoter regions¹⁸.

Table 1 | Chromosome distribution for the 158 IQ-related genes

Chromosome	Number of Genes	Proportion
ChrX*	25	0.1582
Chr7	14	0.0886
Chr17	11	0.0696
Chr1	10	0.0633
Chr12	9	0.0570
Chr11	9	0.0570
Chr9	8	0.0506
Chr6	8	0.0506
Chr8	7	0.0443
Chr3	7	0.0443
Chr5	6	0.0380
Chr20	6	0.0380
Chr2	6	0.0380
Chr15	6	0.0380
Chr16	5	0.0316
Chr4	4	0.0253
Chr19	4	0.0253
Chr13	4	0.0253
Chr22	3	0.0253
Chr21	3	0.0253
Chr10	2	0.0190
ChrY*	1	0.0063
Chr14	1	0.0063

Note: * gene GTPBP6 is located in both X and Y chromosomes.

The other major group of regulators, MicroRNA (miRNA), can regulate mRNA expression at the post-transcriptional level, such as degradation or translational repression by binding the target gene with small complementary sequences¹⁹. To identify potential upstream TFs and miRNAs as possible regulators of the 158 gene set, we performed enrichment analysis on both TF and miRNA targets. In total, we identified seven TFs likely to regulate the set of IQ-related genes. They are FOXF2, FOXO4, MAZ, MEF2A, NFIL3, TCF3, and TFAP4. NFIL3 is reportedly related to neuron disease²⁰. Most remarkably, MEF2 is demonstrated to negatively regulate learning-induced structural plasticity and memory formation²¹. There are 16 target genes for MEF2 in 158 gene set, including *ATXN1*, *BDNF*, *BRAF*, *DBH*, *DMD*, *DMPK*, *DRD3*, *GNAS*, *GRIN2B*, *IL1RAPL1*, *IL6R*, *NR3C2*, *PHOX2B*, *SNAP25*, *TMEM67*, and *TSC1*. Based on pathway annotation, we found that six of them (*BDNF*, *BRAF*, *DRD3*, *GNAS*, *GRIN2B*, and *SNAP25*) had functions in the neuronal system. In addition, the genes *DRD3*, *GNAS*, and *GRIN2B* were related to dopaminergic synapse. These MEF2 targets in the IQ-related gene set may form the core transcriptional circuitry influencing memory formation related to IQ, which begs a further experimental validation.

Based on a similar approach, we found that 16 human miRNAs tended to regulate IQ-related genes. They were hsa-let-7i, hsa-mir-15b, hsa-mir-16, hsa-mir-181d, hsa-mir-195, hsa-mir-30a-3p, hsa-mir-30e-3p, hsa-mir-330, hsa-mir-374, hsa-mir-424, hsa-mir-429, hsa-mir-497, hsa-mir-515-5p, hsa-mir-519e, hsa-mir-522, and hsa-mir-96. Four of them, hsa-let-7i, hsa-miR-15b, hsa-miR-195, and hsa-miR-330, were related to mental disorders based on disease annotation. Both hsa-miR-15b and hsa-miR-195 belong to the miR-15 family, which has been up-regulated in the superior temporal gyrus and the dorsolateral prefrontal cortex in schizophrenics²². These two miRNAs can mediate wide gene silence in the cell. Among the 158 IQ-related genes, the targets of these two miRNAs are *ADRB2*, *ATXN2*, *BDNF*, *GHR*, *IL1RAPL1*, *KCNJ2*, *MAP2K1*, *PAFAH1B1*, *RAF1*, *RELN*, *RPS6KA3*, *SIGMAR1*, *SLC6A4*, and *STX1A*. Comparing this set with that of the MEF2 targets, only *BDNF* occurs in both sets of targets of mental-disorder-related TFs and miRNAs, which might highlight its central role in signaling



pathways related to human intelligence and its potential role as the hub of regulatory circuits in the IQ-related genes²³.

Reconstruction of the core pathway for IQ-related genes using known biological pathways. We are now able to specify several genomic regions and regulators as possible determinants of the IQ phenotype. We adopted a pathway reconstruction approach to describe more complex biological processes arranged in the form of a cascade of connected biochemical reactions or signaling transductions. Generally, molecular biological experimental technology enables us to identify physical and functional interactions between molecules in the cell. Plenty of signaling networks and pathways are summarized based on reliable experimental evidence. Generally speaking, the biological pathways are scattered in many databases and are often represented as diagrams. Recently, Pathway Commons integrated popular pathway databases and now provides a convenient plain-text-based convenient format for further pathway mining and reconstruction²⁴. In order to utilize the available data, we adopted the Klein-Ravi Steiner algorithm to extract core interactions from the integrated human pathway data (see Methods). The reconstructed core pathway contains 97 genes and 129 fully connected pathway connections in total (Figure 1A). Among the 97 nodes, 62 are in the 158 IQ-related genes.

Biological pathway enrichment analysis is one of the most practical ways to mine underlying molecular mechanisms in complex

cellular processes²⁵. Further functional enrichment analysis showed that 97 genes in our reconstructed map were enriched in 30 biological pathways (Table 2). In terms of neuron-related function, there are three identified pathways: “Neurotrophin signaling”, “Long-term potentiation”, and “GnRH signaling”. Interestingly, the majority of the remaining 27 pathways are related to various signaling events, including signaling in cancer, ErbB receptor, TRAIL, proteoglycan syndecan, IFN-gamma, PI3K, MAPK, TSH, Kit receptor, TCR, IL-3, and hepatocyte Growth Factor Receptor. Besides neurohormone GnRH, which is produced in a neural cell and released at its neural terminal, the map indicated that five additional hormones may have an effect. These are androgen, endothelins, glypican, leptin, and prolactin. In summary, our reconstructed map revealed multiple paths related to several known signaling pathways, suggesting potential cellular mechanisms which have not been presented on the topic of signaling transduction, as far as we know. Hormone related pathways, including an androgen reception signaling pathway related to development of male secondary sex characteristics, might suggest interesting and new components related to sex difference, extending our current knowledge.

The arrangement and structure of the nodes in a complex system, such as a biological network, often follow specific rules which may be closely related to the function in this system¹¹. To decompose the reconstructed maps, topological analysis was conducted (Figure 1B–D). Generally, the number of connections at each node is represented

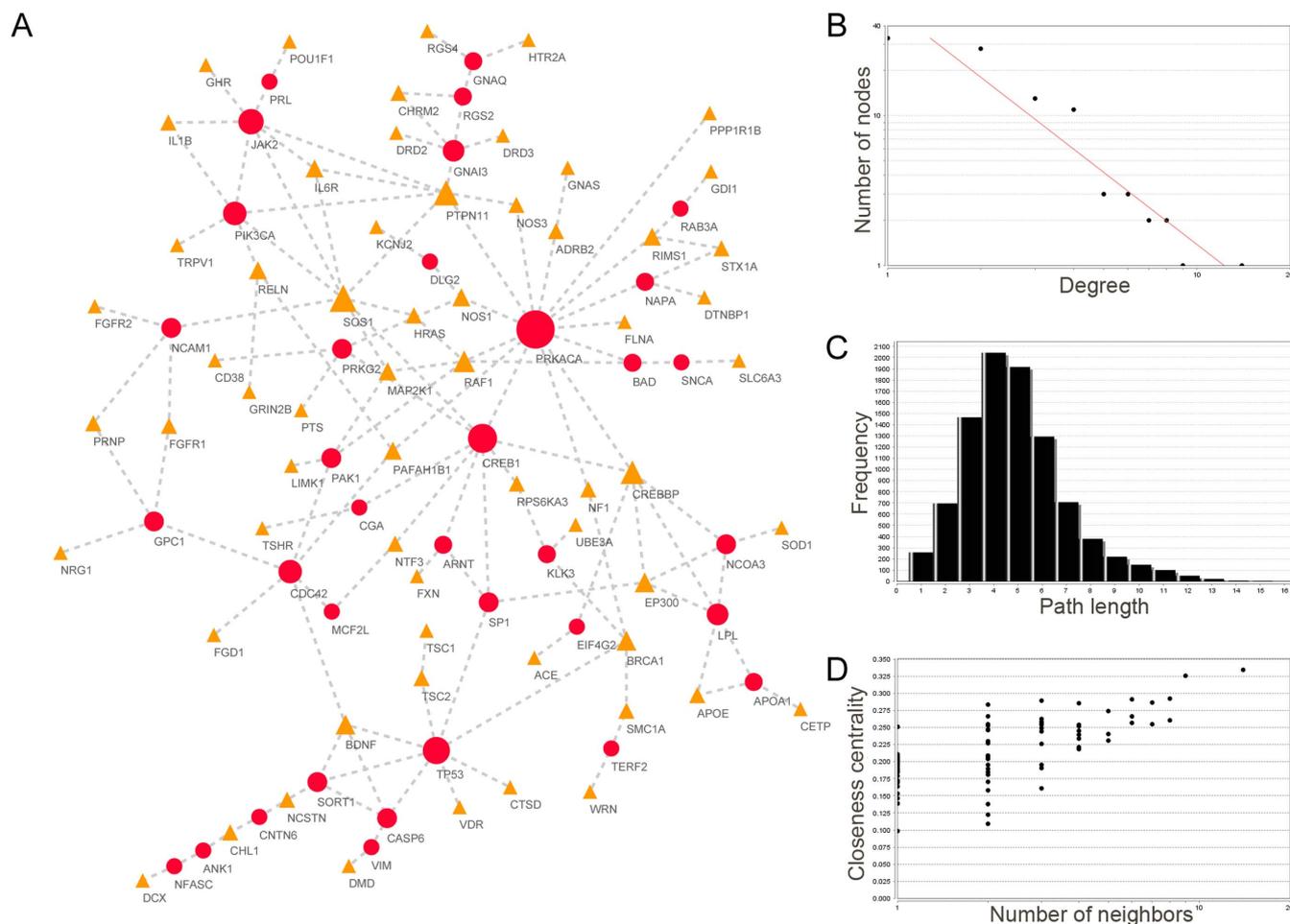


Figure 1 | Network of 158 IQ-related genes based on known pathway interaction data. (A) Reconstructed biological network using 158 IQ-related genes as input by integrating gene-gene interaction from well-defined pathway data. The nodes in yellow (triangle) represent literature-based IQ-related genes; nodes in red (circle) represent expanded genes based on pathway-interaction data. The size of each node represents the number of connection in this network. (B) The plot of degrees and number of nodes in the reconstructed IQ-specific network. (C) The histogram of path length in the reconstructed IQ-specific network. (D) The plot of closeness centralities and the number of neighbors in the reconstructed IQ-specific network.


Table 2 | Significant enriched pathways in the reconstructed IQ-specific network

Pathway	Adjusted P-values*
Pathway Commons pathway	
Glypican pathway	5.23E-39
Glypican 1 network	2.01E-38
ErbB receptor signaling network	5.53E-37
TRAIL signaling pathway	6.90E-37
Proteoglycan syndecan-mediated signaling events	9.40E-37
IFN-gamma pathway	4.30E-36
Plasma membrane estrogen receptor signaling	4.32E-36
Endothelins	4.56E-36
Internalization of ErbB1	2.08E-35
Class I PI3K signaling events	2.08E-35
KEGG pathway	
Prostate cancer	1.74E-18
MAPK signaling pathway	6.92E-17
Neurotrophin signaling pathway	7.31E-17
Renal cell carcinoma	1.75E-16
Gap junction	2.59E-15
Pathways in cancer	3.47E-13
Melanogenesis	4.15E-13
Long-term potentiation	6.36E-13
Regulation of actin cytoskeleton	7.35E-13
GnRH signaling pathway	1.54E-11
Wikipathways pathway	
Prolactin Signaling Pathway	1.30E-15
TSH signaling pathway	1.02E-14
MAPK signaling pathway	5.83E-14
Androgen receptor signaling pathway	8.47E-14
Kit receptor signaling pathway	8.47E-14
TCR Signaling Pathway	1.37E-13
Leptin signaling pathway	1.39E-13
Regulation of Actin Cytoskeleton	3.44E-13
IL-3 Signaling Pathway	7.85E-13
Signaling of Hepatocyte Growth Factor Receptor	6.85E-12

Note: *Adjusted P-values: the P-values of the hypergeometric test were corrected by Benjamini-Hochberg multiple testing correction.

as the degree in a network¹¹. As shown in Figure 1B, the degrees of all molecules in the reconstructed map follow a power law distribution: $P(k) \sim k^{-b}$, where $P(k)$ is the probability that a molecule connects with k molecules and b has an estimated value of 1.602. Therefore the majority of molecules in our map are sparsely connected. In contrast, a small fraction of molecules are more likely to be connected. In total, there are 11 molecules with at least five connections. They are *PRKACA* (14), *CREB1* (9), *TP53* (8), *SOS1* (8), *JAK2* (7), *PTPN11* (7), *PIK3CA* (6), *CREBBP* (6), *CDC42* (6), *RAF1* (5), and *GNAI3* (5). Among these, four are in 158 IQ-related gene set, namely *SOS1*, *PTPN11*, *CREBBP*, and *RAF1*. The remaining seven molecules are appended through pathways connected to the 158 genes.

The hub nodes in a network often serve as common connections to mediate information transduction along a short path. Thus, they often play a prominent role in biological network. In our map, gene *PRKACA* is the most connected, showing 14 connections. *PRKACA* (protein kinase, cAMP-dependent, catalytic, alpha) plays a fundamental role in various cellular functions related to 76 KEGG and 59 REACTOME pathways including cell cycle, apoptosis, signaling transduction, gap junction, and interaction with HIV and the immune system. In addition, *PRKACA* involves many neurological processes such as Long-term potentiation, GnRH signaling, Nicotine Activity on Dopaminergic Neurons, addiction (Amphetamine, Cocaine, and Morphine), synapse activities (Cholinergic, Dopaminergic, Glutamatergic and Serotonergic synapse). Previous studies demonstrates that cAMP-dependent protein kinases are involved in the associative learning of the *Drosophila* (fruit fly)²⁶.

Additionally, cAMP/cAMP-dependent protein kinases in the hippocampal region are reported to be related to a late memory consolidation phase of aversively motivated learning in rats²⁷. Moreover, cAMP-dependent protein kinases can also cooperate with CaMKII in the H3 receptor to regulate the synthesis and release of histamine²⁸. In addition to *PRKACA*, *CREB1* (cAMP response element-binding protein 1), the second-most connected gene in our reconstructed map, can also bond to cAMP response elements in DNA. This gene has been reported to facilitate the formation of long-term memory. Furthermore, *CREB1* interacts with BDNF and NTRK2 to form a core pathway in depression²⁹. In spite of this volume of evidence for the role of cAMP-dependent protein kinases and cAMP response element-binding protein in cognitive-related process, there has been no association between cAMP-dependent protein kinases or cAMP response element-binding protein with IQ. These overlaps with many known signaling cascades with IQ-related genes, which might provide a clue for complex signaling cross-talk centered by *PRKACA/CREB1*, which is in IQdb.

In addition to the cAMP related molecules, there are at least two cancer related genes identified by our pathway reconstruction approach. One is the most well-known tumor suppressor TP53. In fact, there is evidence that it plays a role in IQ-related mental disorders such as schizophrenia³⁰. Furthermore, the other oncogene, *PIK3CA*, is also included in our final map, and is known to influence several psychiatric processes^{31–33}. The next three most-connected genes (*JAK2*, *CDC42*, and *GNAI3*) have also been reported to be related to cognitive disorders or related neuronal functions. *JAK2* is related to cognitive impairment in the mouse model³⁴. *CDC42* is associated with neurofibromatosis and mental retardation³⁵. In spite of the fact that there is no direct evidence for its role in cognitive process, *GNAI3* takes part in negative regulation of synapse transmission, long-term depression, and axon guidance according to KEGG pathway annotation. In summary, of the seven hub molecules in our reconstructed map related to IQ, at least six (85.71%) of them are reported in the literature to be potentially associated with IQ or other cognitive processes. This high relevance not only demonstrates the accuracy of pathway-based reconstruction approach to identify critical molecules, but also provides a fully-connected signaling pathway worthy of further investigation.

Though our reconstructed map is an unoriented signaling pathway, it provides many testable molecules in a typical small-world network involved in cognitive processes, where their degrees follow a power-law distribution. Our further topologic analysis on short path (Figure 1C) and closeness centrality (Figure 1D) shows that the reconstructed map is relatively compact. The path length represents how many steps are between one node and other nodes¹¹. As shown in Figure 1C, the majority of nodes in the map are easily reached from another node in three to six steps. The closeness centrality is used to reveal the shortest step from one node to another¹¹. As shown in Figure 1D, the nodes with more neighbors tend to have higher closeness centrality.

A connected drug-target network in the core pathway of IQ-related genes. Based on the reconstructed pathway map, we further identified enriched drug targets in this map. Using the enriched drug targets as inputs, we combined the drugs and their targets to form a drug-target network. As shown in Figure 2A, there are ten enriched drugs, namely dopamine, nitric oxide, L-tyrosine, methamphetamine, norepinephrine, glutathione, amphetamine, tetrahydrobiopterin, apomorphine, and somatropin recombinant. Except for somatropin recombinant, all have been reported to be active in nervous systems or mental functions according to pharmacodynamics annotation in DrugBank³⁶. Interestingly, six of them interact with two critical neurotransmitters (dopamine and norepinephrine) systems that regulate mood and behavior. These include amphetamine, apomorphine, dopamine, L-tyrosine, and norepinephrine. Dopamine is a neurotransmitter

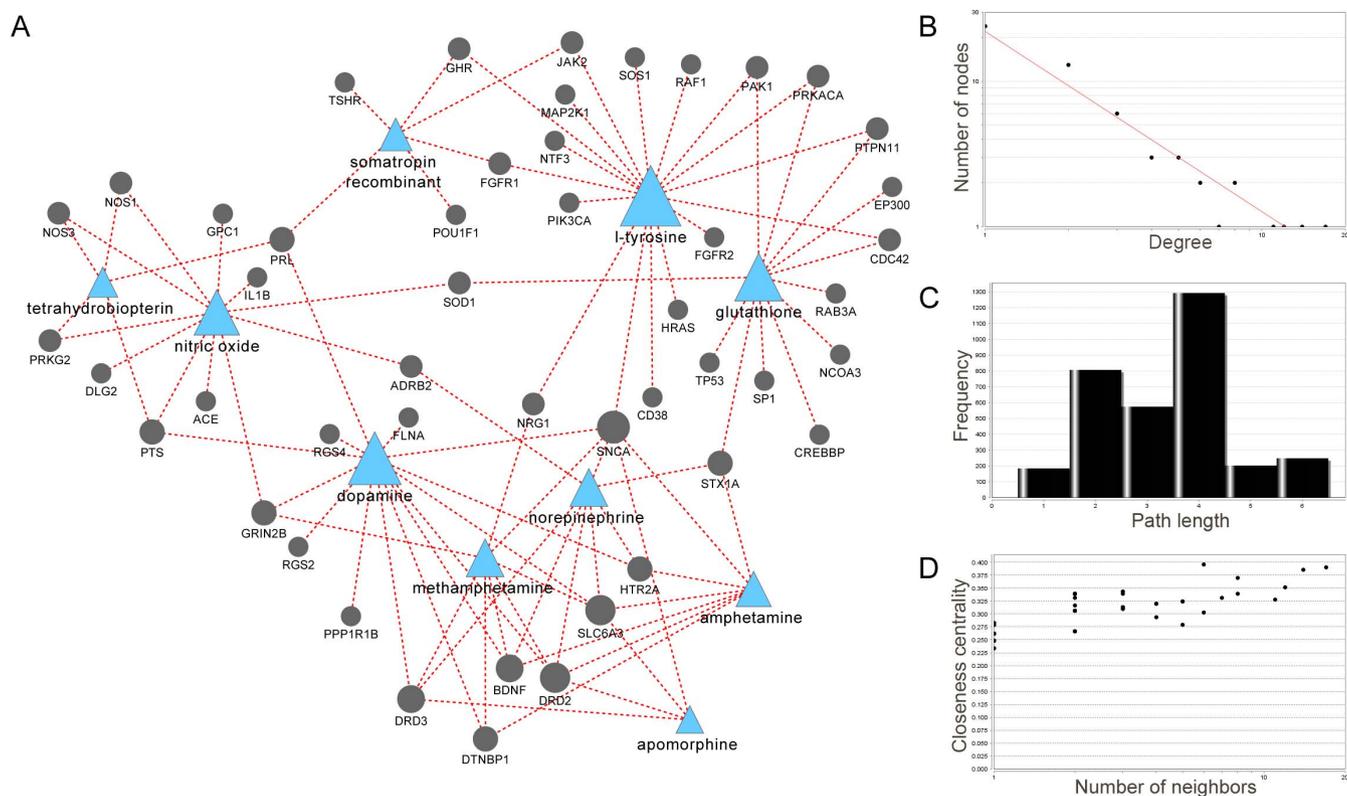


Figure 2 | Network view of drug-target interaction based on reconstructed IQ-specific network. (A) Drug-gene interaction network of IQ-related genes from our reconstructed pathway (see Figure 1). The nodes in blue (triangle) represent the enriched drugs in the reconstructed IQ-specific network (Figure 1); nodes in grey (circle) represent the target of the drugs. The links between drugs and genes represent the drug-target relationships. The size of each node represents the number of connection in this network. (B) The plot of degrees and number of nodes in the IQ-related gene-drug interaction network. (C) The histogram of path length in the IQ-related gene-drug interaction network. (D) The plot of closeness centralities and the number of neighbors in the IQ-related gene-drug interaction network.

responsible for various behavior and cognition activities, such as those found in reward-driven learning systems. Dopamine is also linked to many neurological disorders such as Parkinson's disease, psychosis, and schizophrenia³⁷. Norepinephrine plays a critical role in decision making and can affect attention; it is also used as antidepressant and anti-schizophrenic^{38,39}. L-tyrosine is one of the precursors in the synthesis of dopamine and norepinephrine and can be used to treat depression, improve memory, and enhance mental alertness according to DrugBank annotation³⁶. Amphetamine can stimulate central adrenergic receptors to release norepinephrine and a high dosage of amphetamine is reported to help release dopamine. Methamphetamine, a related entity, is neurotoxic to dopamine transporters and is often used to mark dopamine terminal laboratory animals. Since it is involved in dopamine systems, methamphetamine has been reported to be associated with slower motor function and memory deterioration⁴⁰. Apomorphine is a dopamine agonist and is used to treat Parkinson's disease based on DrugBank annotation, according to DrugBank annotation³⁶.

The six drugs that affect the two critical neurotransmitters dopamine and norepinephrine covered 66.7% of the target genes (32 genes) in the drug-target map (Figure 2A). Since we constructed this map through enriched drugs, this drug-target map is mostly centered by drugs. Therefore, the connectivities of drugs are higher than drug target genes as shown in Figure 2B. Using topological analysis, most of the nodes can be reached from another in two to four steps (Figure 2C). Since the majority of targets and drugs are related to dopamine and norepinephrine system, this network is also very compact based on closeness centralities (Figure 2D). In summary, our enrichment and network-based analysis shows that the dopamine and norepinephrine systems are critical to IQ-related genes,

which may provide more insight into the cognitive process from an IQ aspect.

Discussion

Systems biology, herein defined as the systematic study of complex regulation and interaction in biological systems, makes efforts to analyze cellular processes from a systematic rather than reductionistic perspective, expecting to generate hypotheses to understand the systems. While plenty of single gene-centric studies have been reported as a genetic basis for the explaining IQ, the outcome of the system cannot be fully understood using independent analysis of the parts. Here, we applied an integrative systems-based approach to mine the enriched genomics regions and reconstruct critical biological gene interaction map relevant to IQ. The results support both previously known and novel gene networks related to IQ. For example, we uncovered importance of dopamine and norepinephrine systems in IQ-related biological process. It is still remain unclear about the genetic background about gender-related cognitive processes. Based on functional enrichment analysis, we found that the IQ-related genes were predominantly occurred on X chromosome. This abnormal distribution may highlight the highly rational way to interpret IQ differences between genders. Our integrative analysis may advance the understanding the genomics/cellular factors that promote human cognitive abilities related to IQ, which may elucidate novel pathways for future drug development on IQ-related mental disorders.

We note that the completeness of the cellular processes related to human intelligence is limited by an ever-expanding knowledge of associations between mental phenotype and candidate genes. The associations used here are based on our literature-based database



which currently represents the best available depository of IQ-related genes. It continues to develop quickly due to high-throughput technologies for identifying IQ-related gene associations. Such expansion might unveil the complex signaling cross-talks of human intelligence. Significant improvements are expected from more accurate reconstructions of the human intelligence pathway map, from the more comprehensive integration of various new static and dynamic data, and from improvements in network-based pathway reconstruction approaches. Although the approach may be promising, multiple testable results in our study will not significantly change, i.e. the finding that cAMP related cellular events in the IQ-related pathway highlight the importance of cAMP signaling. In addition, a brain-specific miR-134, localizing hippocampal dendritic spines, is involved in regulation of dendritic spine size²². This synaptic change is related to human memory and plasticity. Though our miRNA target analysis did not intersect with miR-134, we identified four miRNAs related to mental disorders; two of them belonging to the miR-15 family²².

In our reconstructed pathway map, numerous genes were involved in tumorigenesis. Our observation that the oncogenic genes tend to occur in the IQ-related cellular process raises an important question: Is there a common pathway between cancer cell proliferation and the IQ-related cellular process? Previous studies show that there is an extensive connection between cancer and neurodegeneration disorders. For example, the defects of cell-cycle and DNA damage/repair genes both lead to tumorigenesis and neuron diseases⁴¹. One of most well-known Parkinson's causal gene *PARKIN* has been reported as a tumor suppressors in several cancer types⁴². Though there is no direct evidence to associate cognitive processes with cancer, our previous pathway analysis showed that IQ-related genes were highly related to neuron disorders such as schizophrenia, autism, and bipolar⁷. Therefore, it is reasonable to expect numerous overlaps in molecular pathways implicated in cancer and IQ-related cellular processes. We showed that the targets of two schizophrenia-related miRNAs are enriched in our 158 IQ-related genes. Both of these are from miR-15 family, which has been identified as tumor suppressors targeting oncogenes *BCL2*^{43,44}. In summary, the evidence may provide an opportunity to reuse known therapeutic approaches in both groups of disorders.

A fundamental goal of biomedical research is to discover effective drugs to cure human diseases. Our rudimentary analysis of a potential drug target network allows us to relate the two critical neurotransmitters dopamine and norepinephrine systems to IQ-scores. Though dopamine is known to be important in reward-driven learning systems, its broader effect on human intelligence has not been explored. In addition, our constructed drug-target network implied that dopamine and norepinephrine may cooperate with each other to influence other common targets related to IQ. Taken together, these results offer insight into the functional importance of several neurotransmitters in human cognitive process and their relevance to human mental health.

Our systematic analysis of IQ-related genes from the literature resulted in numerous testable hypotheses regarding critical genomic regions, transcription factors, miRNAs, signaling molecules, and neurotransmitters. Most importantly, we found that the systematic mapping of these genes between various annotations helped us uncover some critical biological events and helped to explain their underlying interplay in the human cognitive process.

Methods

Collecting IQ-related genes from the literature. To systematically study the IQ-related genes, we performed an extensive literature search followed by manual assembly of the data. Our literature search focused on family-based linkage studies, population association studies, genome-wide association studies and other functional analyses using the following expression against PubMed database: (“intelligence quotient” [Title/Abstract] OR “IQ” [Title/Abstract]) AND (“genome-wide association study” [Title/Abstract] OR “genome wide association study” [Title/Abstract]) OR (“gene” [Title/Abstract] OR “genetic” [Title/Abstract]) OR

(“association” [Title/Abstract] OR “linkage” [Title/Abstract])). This search returned 2307 abstracts on 10 January 2013. We manually assembled the experimentally verified candidate genes, single-nucleotide polymorphisms (SNPs), and genomic regions relevant to IQ. This information and related annotation were stored in the IQdb database and published for public use⁷. In this article, all 158 literature-based and IQ-related genes were downloaded from IQdb website⁷.

Biological function enrichment analysis. To assess the function of all the 158 IQ-related genes, we conducted functional enrichment tests using the online tool WebGestalt⁴⁵. WebGestalt adopts a hypergeometric model to measure whether an input set of object pairs has a different frequency of annotation pairs than would occur randomly. To study the genomics features of the 158 IQ-related genes, we conducted chromosome cytoband-based enrichment analysis to identify the genomic regions which IQ-related genes were significantly enriched using all the genes on these regions as background. Similar processes were used to identify enriched TF targets, miRNA targets, and drug targets. The identified TFs and miRNAs were regarded as the potential upstream regulators for the IQ-related genes. In these enrichment analyses, all the human protein-coding genes in WebGestalt were used as background to calculate statistical significance. In addition, the Benjamini-Hochberg method was implemented in the WebGestalt to further exclude false negative results. Finally, we identified those enriched functional terms with an adjusted P-value less than 0.05, as calculated by the hypergeometric test.

Network reconstruction, visualization, and topologic analysis. To reconstruct a network based on the 158 IQ-related genes, we first downloaded a non-redundant human pathway interaction from Pathway Commons Database²⁴. In Pathway Commons, it integrates 1,668 pathways from public pathway databases, including KEGG⁴⁶, BioCyc⁴⁷, Reactome⁴⁸, and pathway interaction database⁴⁹. Based on the reactions in these pathway maps, Pathway Commons decomposes these pathways into pairs of protein-protein interactions. This pathway interaction data can be used to reconstruct biological pathways efficiently. In total, the final human pathway-based interactome contained 3629 nodes and 36034 pathway interaction links. To build a non-redundant protein-protein interaction dataset, all the self-interaction were excluded. Using the Klein-Ravi Steiner minimal tree algorithm⁵⁰, we reconstructed a sub-network as an IQ-specific pathway from the entire human pathway-based interactome. Among the 158 IQ-related genes, only 112 were mapped to data from Pathway Commons.

Visualization of the reconstructed pathway network was performed using the Cytoscape software⁵¹. The input to Cytoscape is tab delimited file with two columns with two interacting genes. The node is corresponding to each interactors, the edge denotes the pathway interaction between the two interacting genes. Node shapes and colors were used to represent molecules with different functions. Links between any two nodes indicated their functional relevance.

To obtain the network topological properties, we calculated the number of connections (degree), the path length, and the closeness centrality for each node in the reconstructed IQ core pathway using NetworkAnalyzer toolkit in Cytoscape⁵¹. The degree is a local property to represent the number of connections of each node. The higher the degree, the higher the probability of this gene which interacts to other genes in the reconstructed network. The path length is defined as the number of steps along the shortest paths from a specific node to the remaining network nodes. The shortest path length distribution gives the number of node pairs with different number of step to communicate. The closeness centrality calculates how many steps a specific node requires to connect to other nodes. Closeness centrality is preferred to mean shortest-path length, as it gives higher values to more central nodes, and so is usually positively correlated to degree.

1. Neisser, U. *et al.* Intelligence: Knowns and unknowns. *Am Psychol* **51**, 77–101 (1996).
2. Kremen, W. S. *et al.* Heritability of word recognition in middle-aged men varies as a function of parental education. *Behav Genet* **35**, 417–433 (2005).
3. Deary, I. J., Johnson, W. & Houlihan, L. M. Genetic foundations of human intelligence. *Hum Genet* **126**, 215–232 (2009).
4. Nisbett, R. E. *et al.* Intelligence New Findings and Theoretical Developments. *Am Psychol* **67**, 130–159 (2012).
5. Plomin, R. & Haworth, C. M. A. Genetics of High Cognitive Abilities. *Behav Genet* **39**, 347–349 (2009).
6. Plomin, R. & Spinath, F. M. Intelligence: genetics, genes, and genomics. *J Pers Soc Psychol* **86**, 112–129 (2004).
7. Kong, L., Fan, L., Zhao, M. & Qu, H. IQdb: an intelligence quotient score associated gene resource for human intelligence. *Database (Oxford)* **2013**, bat063 (2013).
8. Mortensen, E. L., Sorensen, H. J., Jensen, H. H., Reinisch, J. M. & Mednick, S. A. IQ and mental disorder in young men. *Br J Psychiatry* **187**, 407–415 (2005).
9. Koenen, K. C. *et al.* Childhood IQ and adult mental disorders: a test of the cognitive reserve hypothesis. *Am J Psychiatry* **166**, 50–57 (2009).
10. Kalbfleisch, M. L. & Loughan, A. R. Impact of IQ Discrepancy on Executive Function in High-Functioning Autism: Insight into Twice Exceptionality. *J Autism Dev Disord* **42**, 390–400 (2011).
11. Barabasi, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**, 101–113 (2004).



12. Skuse, D. H. X-linked genes and mental functioning. *Hum Mol Genet* **14 Spec No 1**, R27–32 (2005).
13. Laumonier, F., Cuthbert, P. C. & Grant, S. G. The role of neuronal complexes in human X-linked brain diseases. *Am J Hum Genet* **80**, 205–220 (2007).
14. Blinkhorn, S. Is there a sex difference in IQ scores? Reply. *Nature* **442**, E1–E2 (2006).
15. Irwing, P. & Lynn, R. Is there a sex difference in IQ scores? *Nature* **442**, E1–E1 (2006).
16. Wright, M. J. G. N., Luciano, M., Zhu, G. & Martin, N. G. [Genetics of personality and cognition in adolescents.] *Developmental psychopathology and wellness: genetic and environmental influences, 1st edn* [85–108] (American Psychiatric Publishing Inc, 2008).
17. Luciano, M. *et al.* Genome-wide scan of IQ finds significant linkage to a quantitative trait locus on 2q. *Behav Genet* **36**, 45–55 (2006).
18. Darnell, J. E. Transcription factors as targets for cancer therapy. *Nat Rev Cancer* **2**, 740–749 (2002).
19. Farazi, T. A., Spitzer, J. I., Morozov, P. & Tuschl, T. miRNAs in human cancer. *J Pathol* **223**, 102–115 (2011).
20. Hulme, D. J., Blair, I. P., Dawkins, J. L. & Nicholson, G. A. Exclusion of NFIL3 as the gene causing hereditary sensory neuropathy type I by mutation analysis. *Hum Genet* **106**, 594–596 (2000).
21. Cole, C. J. *et al.* MEF2 negatively regulates learning-induced structural plasticity and memory formation. *Nat Neurosci* **15**, 1255–1264 (2012).
22. Beveridge, N. J., Gardiner, E., Carroll, A. P., Tooney, P. A. & Cairns, M. J. Schizophrenia is associated with an increase in cortical microRNA biogenesis. *Mol Psychiatry* **15**, 1176–1189 (2010).
23. Tsai, S. J., Hong, C. J., Yu, Y. W. & Chen, T. J. Association study of a brain-derived neurotrophic factor (BDNF) Val66Met polymorphism and personality trait and intelligence in healthy young females. *Neuropsychobiology* **49**, 13–16 (2004).
24. Cerami, E. G. *et al.* Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res* **39**, D685–690 (2011).
25. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1–13 (2009).
26. Drain, P., Folkers, E. & Quinn, W. G. cAMP-dependent protein kinase and the disruption of learning in transgenic flies. *Neuron* **6**, 71–82 (1991).
27. Bernabeu, R. *et al.* Involvement of hippocampal cAMP/cAMP-dependent protein kinase signaling pathways in a late memory consolidation phase of aversively motivated learning in rats. *Proc Natl Acad Sci U S A* **94**, 7041–7046 (1997).
28. Moreno-Delgado, D. *et al.* Different role of cAMP dependent protein kinase and CaMKII in H3 receptor regulation of histamine synthesis and release. *Neuroscience* **164**, 1244–1251 (2009).
29. Juhasz, G. *et al.* The CREB1-BDNF-NTRK2 pathway in depression: multiple gene-cognition-environment interactions. *Biol Psychiatry* **69**, 762–771 (2011).
30. Papiol, S. *et al.* Analysis of polymorphisms at the tumor suppressor gene p53 (TP53) in contributing to the risk for schizophrenia and its associated neurocognitive deficits. *Neurosci Lett* **363**, 78–80 (2004).
31. Abdul, H. M. & Butterfield, D. A. Involvement of PI3K/PKG/ERK1/2 signaling pathways in cortical neurons to trigger protection by cotreatment of acetyl-L-carnitine and alpha-lipoic acid against HNE-mediated oxidative stress and neurotoxicity: implications for Alzheimer's disease. *Free Radic Biol Med* **42**, 371–384 (2007).
32. Sun, R., Yan, J. & Willis, W. D. Activation of protein kinase B/Akt in the periphery contributes to pain behavior induced by capsaicin in rats. *Neuroscience* **144**, 286–294 (2007).
33. Tornieri, K., Welshhans, K., Geddis, M. S. & Rehder, V. Control of neurite outgrowth and growth cone motility by phosphatidylinositol-3-kinase. *Cell Motil Cytoskeleton* **63**, 173–192 (2006).
34. Park, S. J. *et al.* Inactivation of JAK2/STAT3 Signaling Axis and Down-Regulation of M1 mAChR Cause Cognitive Impairment in klotho Mutant Mice, a Genetic Model of Aging. *Neuropsychopharmacology* **38**, 1426–1437 (2013).
35. Node-Langlois, R., Muller, D. & Boda, B. Sequential implication of the mental retardation proteins ARHGEF6 and PAK3 in spine morphogenesis. *J Cell Sci* **119**, 4986–4993 (2006).
36. Knox, C. *et al.* DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res* **39**, D1035–1041 (2011).
37. Birtwistle, J. & Baldwin, D. Role of dopamine in schizophrenia and Parkinson's disease. *Br J Nurs* **7**, 832–834 (1998).
38. Long, L. E., Malone, D. T. & Taylor, D. A. Cannabidiol reverses MK-801-induced disruption of prepulse inhibition in mice. *Neuropsychopharmacology* **31**, 795–803 (2006).
39. De Backer, D. *et al.* Comparison of dopamine and norepinephrine in the treatment of shock. *N Engl J Med* **362**, 779–789 (2010).
40. Volkow, N. D. *et al.* Loss of dopamine transporters in methamphetamine abusers recovers with protracted abstinence. *J Neurosci* **21**, 9414–9418 (2001).
41. Staropoli, J. F. Tumorigenesis and neurodegeneration: two sides of the same coin? *Bioessays* **30**, 719–727 (2008).
42. Garber, K. Parkinson's disease and cancer: the unexplored connection. *J Natl Cancer Inst* **102**, 371–374 (2010).
43. Finnerty, J. R. *et al.* The miR-15/107 group of microRNA genes: evolutionary biology, cellular functions, and roles in human diseases. *J Mol Biol* **402**, 491–509 (2010).
44. Cimmino, A. *et al.* miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A* **102**, 13944–13949 (2005).
45. Zhang, B., Kirov, S. & Snoddy, J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* **33**, W741–748 (2005).
46. Kanehisa, M. *et al.* From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* **34**, D354–357 (2006).
47. Karp, P. D. *et al.* Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res* **33**, 6083–6089 (2005).
48. Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res* **37**, D619–622 (2009).
49. Schaefer, C. F. *et al.* PID: the Pathway Interaction Database. *Nucleic Acids Res* **37**, D674–679 (2009).
50. Klein, P. & Ravi, R. A Nearly Best-Possible Approximation Algorithm for Node-Weighted Steiner Trees. *J Algorithms* **19**, 104–115 (1995).
51. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504 (2003).

Acknowledgments

This work was supported by the National High-tech 863 Program of China (No. 2006AA02A312, No. 2008BA164B01), and the National Natural Science Foundation of China (No. 31171270).

Author contributions

M.Z. and K.L. carried out the analyses and helped write the manuscript. H.Q. conceived of the analysis and helped write the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zhao, M., Kong, L. & Qu, H. A systems biology approach to identify intelligence quotient score-related genomic regions, and pathways relevant to potential therapeutic treatments. *Sci. Rep.* **4**, 4176; DOI:10.1038/srep04176 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0>