Supporting Information

Lee *et al.* 10.1073/pnas.0802208105

SI Text

Datasets. The KEGG database (1) is a depository of the list of all metabolic reactions identified in a generic human cell, together with their enzymes and the genes encoding them. The BiGG reconstruction was obtained by an iterative process of literature search and curation/extension of the former (2), where compound-specific transport reactions were added from the literature, and reactions that were deemed necessary for the completeness of a particular metabolic pathway, but for which genomic or biochemical evidence could not be found, were also added.

We used the Morbid Map from the Online Mendelian Inheritance in Man (OMIM) (3) to uncover the known gene-disease relationships, providing 2,025 disease genes and 3,423 disease phenotypes as of August 2007. The disorders were grouped (4) into 1,437 distinct disorders associated with 2,025 genes. We found 337 (378) diseases associated with metabolic reactions in the KEGG (BiGG) database, classified into 22 disorder classes, 174 (187) being denoted as classical "metabolic diseases" (see *Disease Classification* below and Dataset S1 for details).

The Medicare dataset contains Medicare claims of 13,039,018 hospitalized patients from 1990 to 1993. Up to 10 diagnosed diseases were recorded in ICD-9-CM format at each of 32,341,347 visits of those patients. The set of patients consists of 5,440,490 males (41.7%) and 7,598,528 females (58.3%), and the age is distributed between 65 and 113 with the average 76.5 and the standard deviation 7.5 (see Fig. S6). The average number of diseases diagnosed for a patient is 8.4, and the average number of patients having a given disease is 5,820. Their distributions are shown in Fig. S6.

Disease Classification. The disease states considered in this work were obtained from the OMIM disease list and are associated with metabolic reactions according to the KEGG or BiGG database. In contrast, traditional medical classification of diseases is based on the known pathophysiology of each disorder as well as historical connotations. To separate classical metabolic diseases (e.g., phenylketonuria) from those that are not considered such but that have mutated enzyme(s) involved in their pathogenesis, we used two separate approaches. First, we used the disease classification reported in ref. 4, where all of the OMIM diseases were manually classified into 22 disorder classes, including the classical metabolic disease class (shown in Dataset S1). Next, we have cross-checked all of the metabolic diseases on this list against the ICD-9-CM and ICD-10-CM coding systems to ensure that they indeed are currently classified as such. Note that although ICD-9-CM and ICD-10-CM coding for metabolic diseases differs from each other in some respect, our purpose here was not to decide which coding system is better but to compile a list of metabolic diseases. When a given disease is categorized into several groups by ICD-9-CM/ ICD-10-CM, of which one is the metabolic disease group, we have assigned them to the latter. This derived classification, i.e. whether a disease is considered metabolic or nonmetabolic, is also shown in Dataset S1.

On many occasions, entering a disease name obtained from ref. 4 into the ICD-9-CM and/or ICD-10-CM returned no match at all. In these cases, we examined the entry through Internet and/or literature search and tagged them as metabolic or nonmetabolic according to the following (somewhat arbitrary) criteria:

- If the disease pathology is strictly organ-specific, we classified it as nonmetabolic. For example, if in a given disease only the cornea is affected, we classified it as eye disease. However, if e.g. concomitant organic acidemia or aciduria is also present, the disease was classified into the metabolic disease class.
- Diseases whose names include "resistance to" or "susceptibility to" were considered nonmetabolic.
- If it is known that the enzyme defect is not involved in the pathology of a disease, it is considered nonmetabolic.
- Mitochondrial and peroxisomal enzyme-caused diseases were classified as metabolic diseases. However, if the mutation caused a generic failure of organelle genesis, we did not consider it as a metabolic disease.
- As a result of these two steps, we classified only 202 of the 415 metabolism-related diseases, according to the KEGG or BiGG database, into the metabolic disease class in either of the two medical classification scheme, as shown in Dataset S1. The remaining ones were classified according to the classification scheme of ref. 4. The diseases in Figs. 1, S3, and S4 were color-coded based on this classification.

Constructing the Metabolic Disease Network (MDN). To establish the metabolic reaction-disease associations, we combined the disease–gene association from the OMIM database and the gene–reaction associations from the KEGG and BiGG database. Two metabolic reactions are adjacent if they convert or produce a common compound, and we can expect their fluxes to be correlated. However, some compounds like ATP, NADH, and H_2O (cofactors) are involved in many reactions, and the flux correlation of two of those reactions may be weak. So we assigned a link to adjacent reactions only when the common compound is involved in no more than n_e distinct reactions. We set $n_e = 4$, but other values can be used for n_e without changing our conclusions. A link was then assigned to a pair of diseases in the MDN if the diseases are associated with the same reaction or their associated reactions are linked. The disease–reaction associations are found in Dataset S1, and the adjacency of the metabolic reactions are in Dataset S2 and Dataset S3.

Flux-Coupled Reactions. The flux-coupling analysis was based on the BiGG database, assuming growth in complex medium in which all of the uptake reactions can occur without limitation. We identified 216 directionally coupled and 86 fully coupled pairs of metabolic reactions, as described in the main text and in ref. 5. The gene–reaction associations allowed us to find 1,925 pairs of directionally and 680 fully coupled pairs of genes, and the disease–reaction associations provided 38 directionally coupled and 20 fully coupled diseases pairs.

Gene–Gene Coexpression Analysis. Among ≈25,000 human genes, 1,197 (1,490) were found to be associated with metabolic reactions according to KEGG (BiGG) database. We analyzed the coexpression level of these metabolism-related genes by using microarray data available for 36 normal human tissues (6). Denoting the expression level of each gene in each tissue by x_{it}, where i is the gene index and t is the tissue index running from 1 to N_t (= 36), we calculated the Pearson correlation coefficient (PCC) for each pair of genes, i and j, as

$$PCC_{ij} = \frac{N_t \sum_t x_{it}x_{jt} - \sum_t x_{it} \sum_t x_{jt}}{\sqrt{N_t \sum_t x_{it}^2 - (\sum_t x_{it})^2}\sqrt{N_t \sum_t x_{jt}^2 - (\sum_t x_{jt})^2}}. \quad [1]$$

Genes associated with the same metabolic reaction or adjacent reactions were assigned a metabolic link as diseases. We also

considered pairs of genes whose associated reactions have directionally coupled or fully coupled fluxes. We compared the PCC of the gene expression profiles for all pairs of metabolism-related genes, those connected by metabolic links, and by flux-coupled links (Fig. 2 *b* and *c*).

Prevalence and Comorbidity Index. The comorbidity is defined as the PCC of their occurrence vectors, O_X and O_Y, defined as $O_X = (O_{X1},O_{X2},\ldots O_{XN})$ and $O_Y = (O_{Y1},O_{Y2},\ldots,O_{YN})$, where $O_{Xp} = 1$ (0) if the patient p has the disease X, and N is the total number of patients in the database. We used a hand-curated map between the genetic disorders in the OMIM and ICD-9-CM codes (Dataset S4) to determine the number of patients diagnosed with each disease (incidence) and with each pair of diseases (coincidence) denoted by N_X and N_{XY}, respectively, in the MDN. The prevalence was defined as

$$I_x = \frac{N_x}{N} \qquad [2]$$

and the comorbidity index was defined as

$$C_{XY} = \frac{NN_{XY} - N_X N_Y}{\sqrt{N_X N_Y (N - N_X)(N - N_Y)}}. \qquad [3]$$

The prevalence and comorbidity indices are found in Dataset S4 and Dataset S5, respectively.

Statistical Significance. We calculated the P values under an appropriate null hypothesis to quantitatively characterize the statistical significance of the obtained results.

High value of the average coexpression for connected genes and average comorbidity for connected diseases. As a null hypothesis, we assumed that the average co-expression (comorbidity) for connected metabolism-related genes (diseases) follows a normal distribution with the average m and the standard deviation $\sigma/\sqrt{N_c}$, where m and σ are the average and the standard deviation of the co-expression (comorbidity) for all metabolism-related genes (diseases), respectively, and N_c is the number of pairs of connected metabolism-related genes (diseases). Therefore an average value r of the co-expression (comorbidity) for connected genes (diseases) gives the z value,

$$z_r = \frac{r - m}{\sigma}\sqrt{N_c - 1}, \qquad [4]$$

and the corresponding P value calculated as

$$P = \int_{z_r}^{\infty} dz \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}. \qquad [5]$$

Significance of the Pearson correlation coefficient. We measured the Pearson correlation coefficients (PCCs) between the disease prevalence/mortality and the disease degree, and between the disease comorbidity and the network distance to characterize quantitatively their correlations. The disease comorbidity itself is also a PCC. Under a null hypothesis that there is no correlation between two given data, their PCC follows the Student's t-distribution. The measured value r of PCC corresponds to the t value,

$$t_r = \frac{r}{\sqrt{1 - r^2}}\sqrt{N - 2} \qquad [6]$$

where N is the number of data. Then the P value for this obtained value of t was calculated by using Student's t-distribution.

Random Network Generation. To compare the degree and distance distribution of the MDN with those that would be without the topological features of the MDN, we generated the networks with the same number of nodes connected randomly by the same number of links as the MDN. To be specific, a random network of N nodes and L links was generated by (i) starting with N isolated nodes, (*ii*) randomly choosing two nodes i and j and assigning a link to them if they are not connected yet, and (*iii*) repeating (*ii*) until the total number of links reaches L. Distributions of the degree and network distance were then obtained by averaging them over 1,000 such random networks and are presented in Fig. S2 along with those for the MDN.

Reaction–Tissue Association and the Core of the Human Metabolic Network. It is well known that different pathways are used under different environments and growth conditions, and recently Almaas *et al.* (7) have computed the core set of metabolic networks through the flux balance analysis of microorganisms *Escherichia coli*, *Helicobacter pylori*, and *Saccharomyces cerevisiae*. Unfortunately, the same methodology cannot be applied to human metabolic networks yet because we are lacking the optimization function to apply flux–balance analysis.

Nevertheless, to approach the question whether the perturbed reactions are located in the core or in the periphery, we defined the tissue-associated core/branches of human metabolic networks. It is obvious that different sets of reactions are active in different tissues, and thus, tissue dependency could be viewed as genetic/epigenetic variants of human metabolism, a role analogous to environmental variables for bacterial and unicellular eukaryotic metabolism. Therefore, we used mRNA expression data from 36 human tissues to define the core/branches of the human metabolism. We find that genes encoding enzymes for ≈35% of all metabolic reactions are simultaneously expressed in all 36 tissues, suggesting that these reactions represent the tissue-independent core of the human metabolism, whereas the rest represent tissue-specific metabolic activities.

Then we investigated the distribution of the number of tissues in which two reactions are active together (see Fig. S13 and Fig. S14 for the metabolic network based on the KEGG and BiGG databases, respectively). We find that most pairs (≈84% for KEGG, ≈82% for BiGG) of reactions are active together in more than one tissue. On the other hand, 11% (12%) of the reaction pairs in the metabolic network from KEGG (BiGG) are active together only in one branch, a tissue specificity that can help drug discovery. Approximately 5% (6%) of the reaction pairs are active in different tissues. We also find that this trend, that two reactions are active together in more than one tissue, is stronger for reactions associated with diseases and even stronger for reactions for which the associated diseases have significant comorbidity ($P < 0.01$). In particular, the fraction of the reaction pairs active together in more than one tissue is 85.5% (86.4%) for disease reaction pairs and 91.1% (86.5%) for comorbid disease reaction pairs in the metabolic network from the KEGG (BiGG) database.

1. Kanehisa M, *et al.* (2006) From genomics to chemical genomics. *Nucleic Acids Res* 34:D354–D357.
2. Duarte ND, *et al.* (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA* 104:1777–1782.
3. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders *Nucleic Acids Res* 33:D514–D517.
4. Goh, K-I, *et al.* (2007) The human disease network. *Proc Natl Acad Sci USA* 104:8685–8690.
5. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD (2004) Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res* 14:301–312.
6. Ge X, *et al.* (2005) Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. *Genomics* 86:127–141.
7. Almaas E, Oltvai ZN, Barabási, A-L (2005) The activity reaction core and plasticity of metabolic networks. *PLoS Comp Biol* 1:0557–0563.

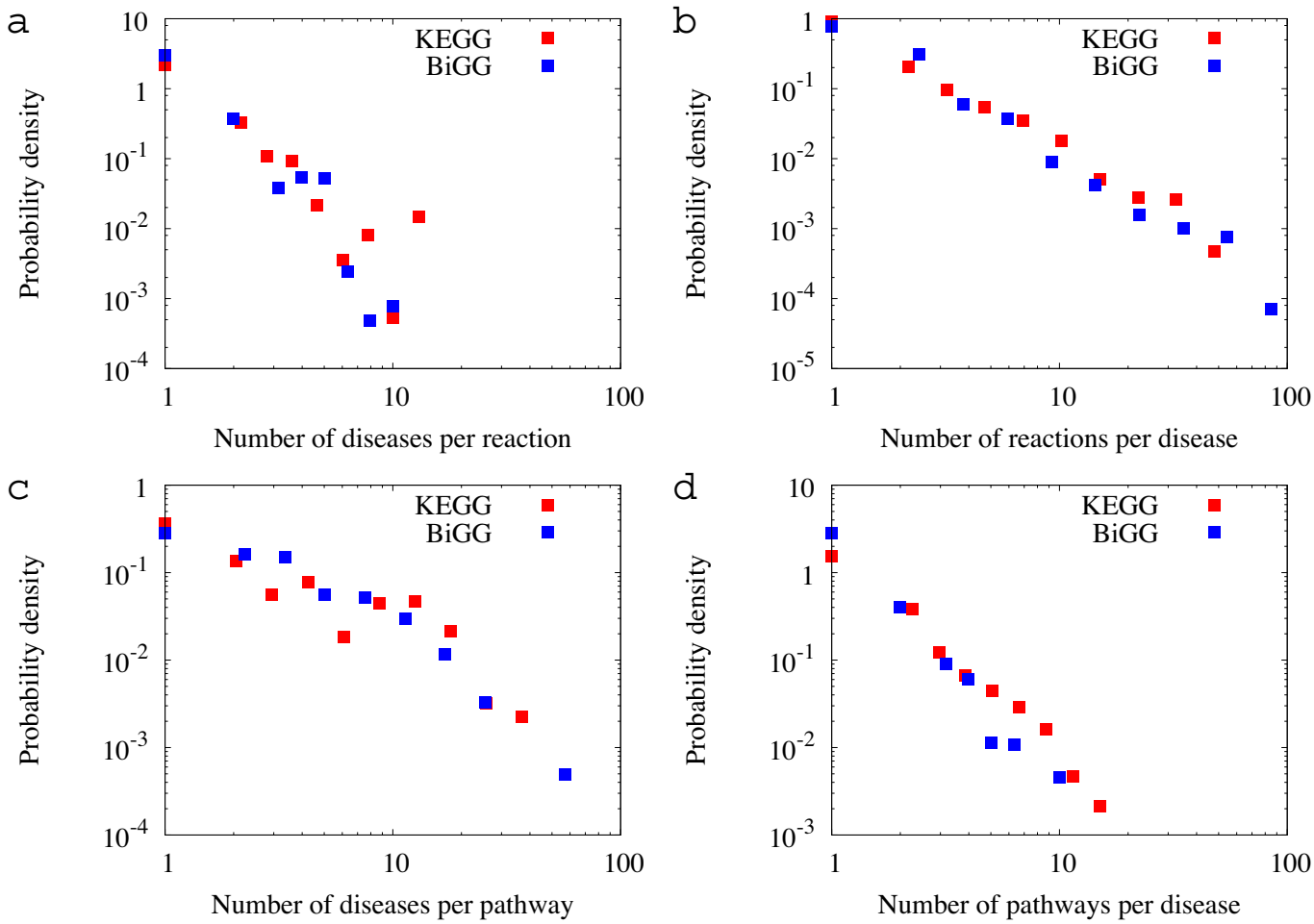


Fig. S1. Statistical characteristics of disease–reaction association. (*a*) Distribution of the number of diseases associated with a reaction. (*b*) Distribution of the number of reactions associated with a disease. (*c*) Distribution of the number of diseases associated with a metabolic pathway defined in KEGG or BiGG database. (*d*) Distribution of the number of pathways associated with a disease. All of the distributions are log-binned.

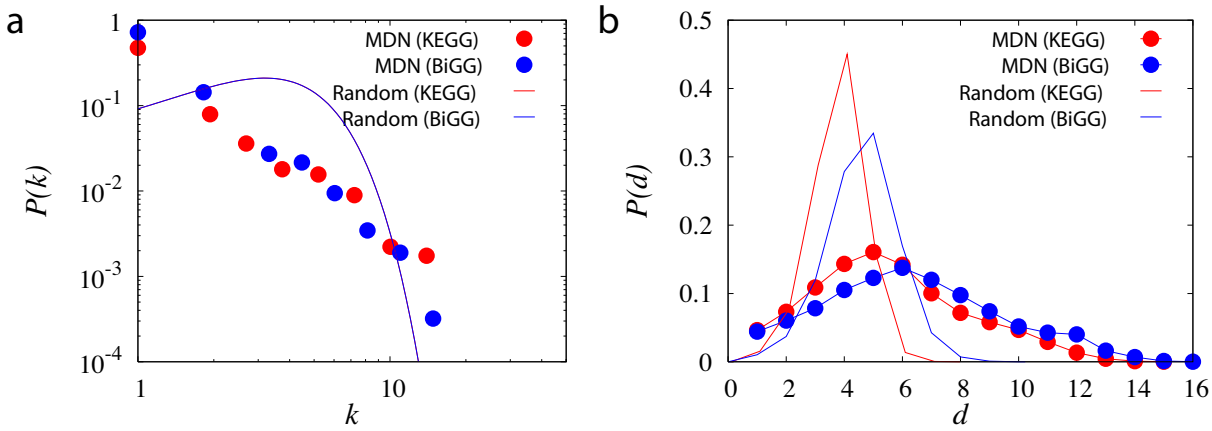


Fig. S2. Distributions of degree and distance in the MDN. (*a*) Degree distribution of the MDN is shown together with that of random networks with the same numbers of nodes and links as the MDN. (*b*) Distance distribution of the MDN is shown together with that of the random networks.

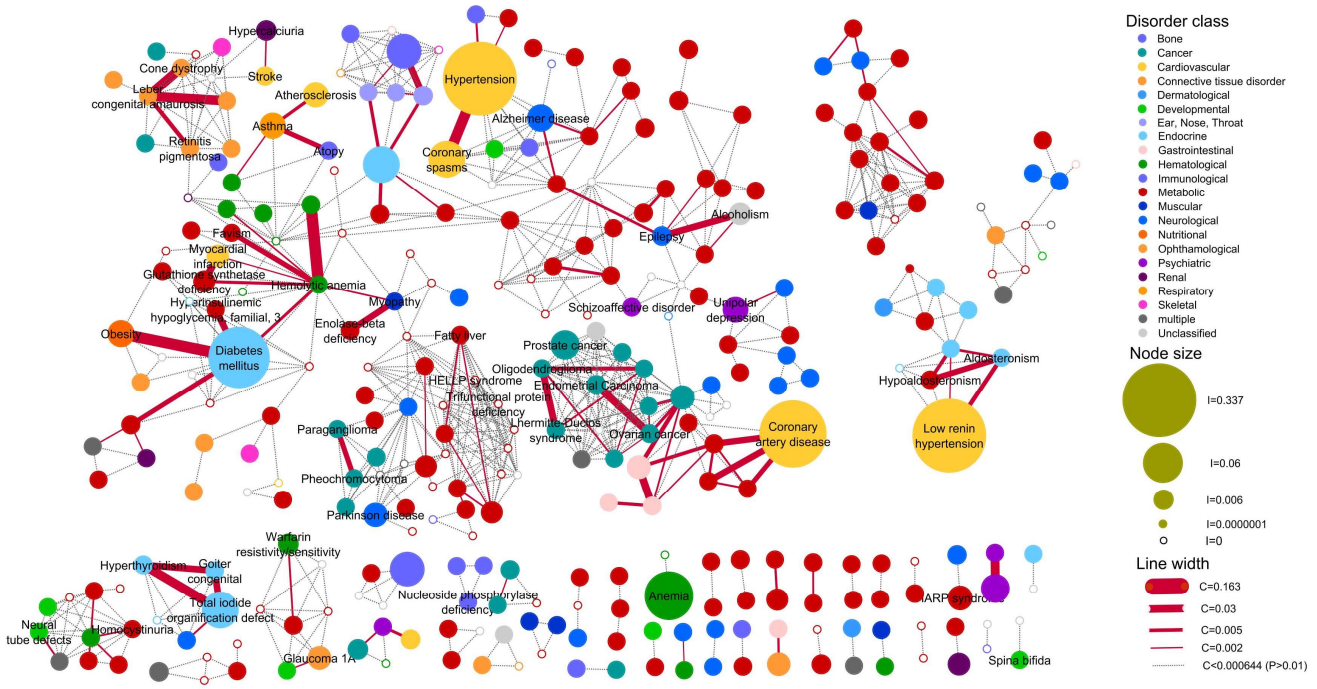


Fig. S3. Metabolic disease network based on the BiGG database. Three hundred and nineteen nonisolated diseases are connected by 699 metabolic links predicted by BiGG and OMIM reconstructions. The color of a node indicates the disorder class it belongs to. Node size is proportional to the prevalence of each disease, and the width of a link is proportional to the comorbidity index of the two diseases the link connects. We show with red the links with significant ($P <$ 0.01) comorbidity.

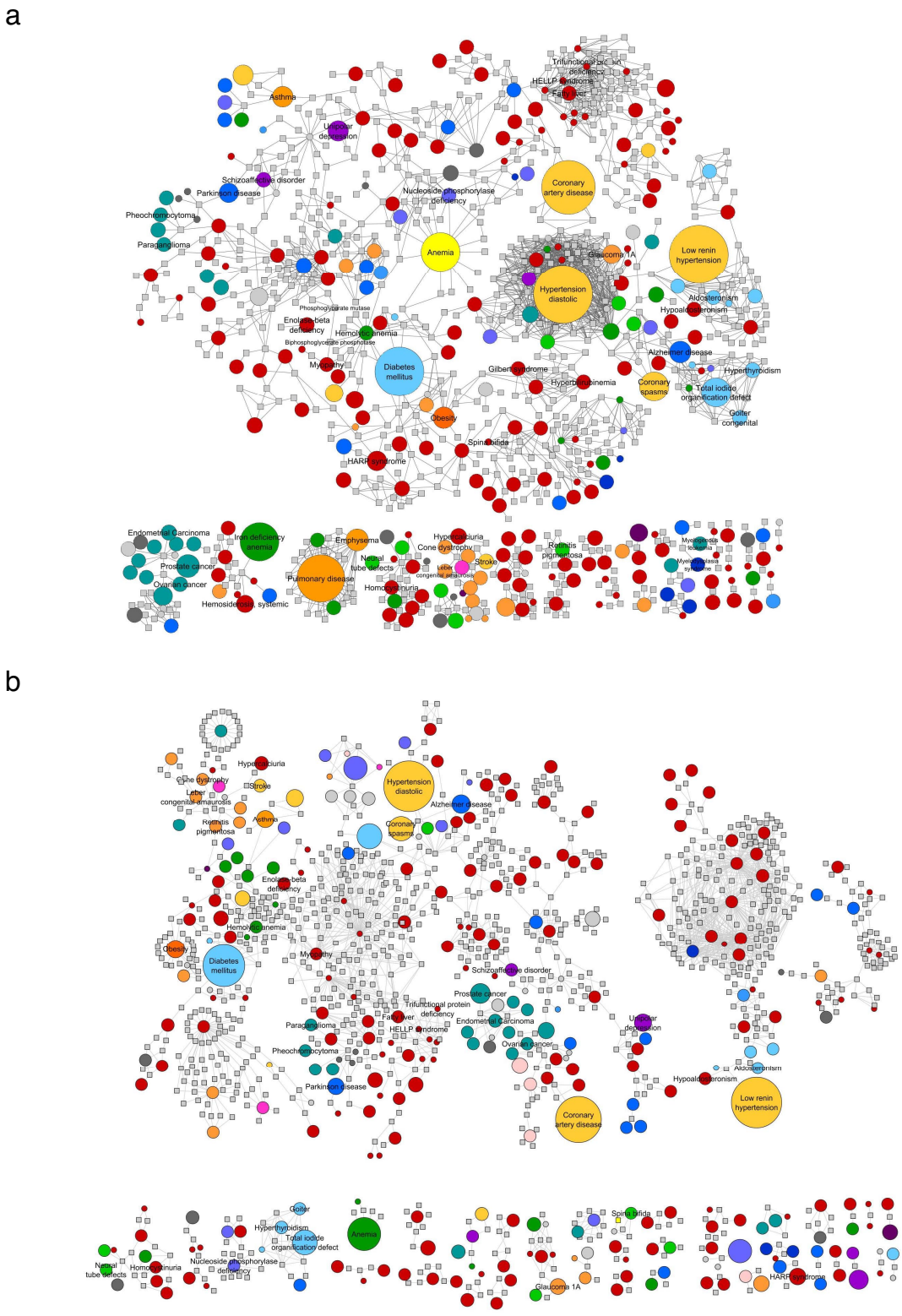


Fig. S4. Disease–reaction bipartite network. The network includes all of the metabolism-related diseases (circle) and the metabolic reactions (square) associated with at least one disease based on the OMIM database and KEGG (*a*) and BiGG (*b*) database. The links are two types: between a disease and a reaction (disease–reaction association) and between two reactions (adjacency). With clusters of only one disease excluded, the network is comprised of 308 (319) diseases, 686 (943) reactions, 1,741 (1,677) disease–reaction links, and 431 (628) reaction–reaction links for KEGG (BiGG) database. The color and size of the diseases nodes reflect the disease class and the prevalence as in the legends of Fig. 1 or Fig. S3.

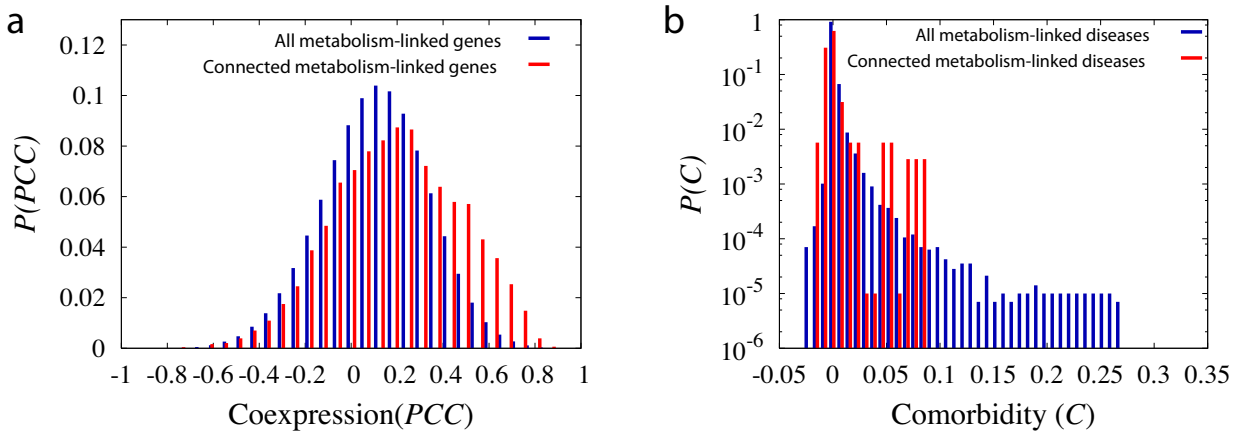


Fig. S5. Gene coexpression and disease comorbidity in the MDN based on the BiGG database. Shown are the distributions of the gene coexpression (*a*) and disease comorbidity (*b*) in the MDN based on the BiGG database.

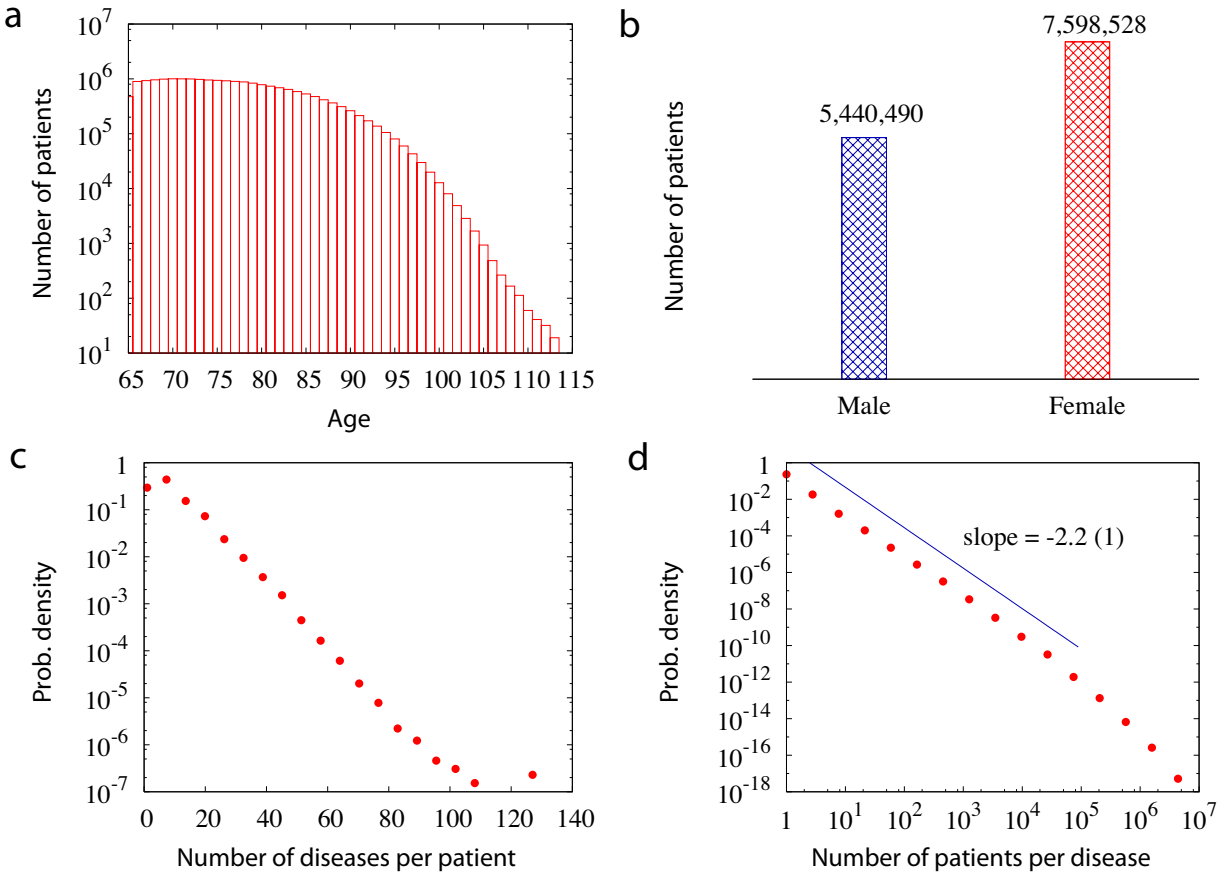


Fig. S6. Characteristics of the Medicare claim dataset. (*a*) Plot of the number of patients versus their age. (*b*) Number of male and female patients. (*c*) Probability density (Prob. density) function of the number of diseases diagnosed for a patient. The *y* axis is in logarithmic scale. (*d*) Probability density function of the number of patients having a disease. Both axes are in logarithmic scale.

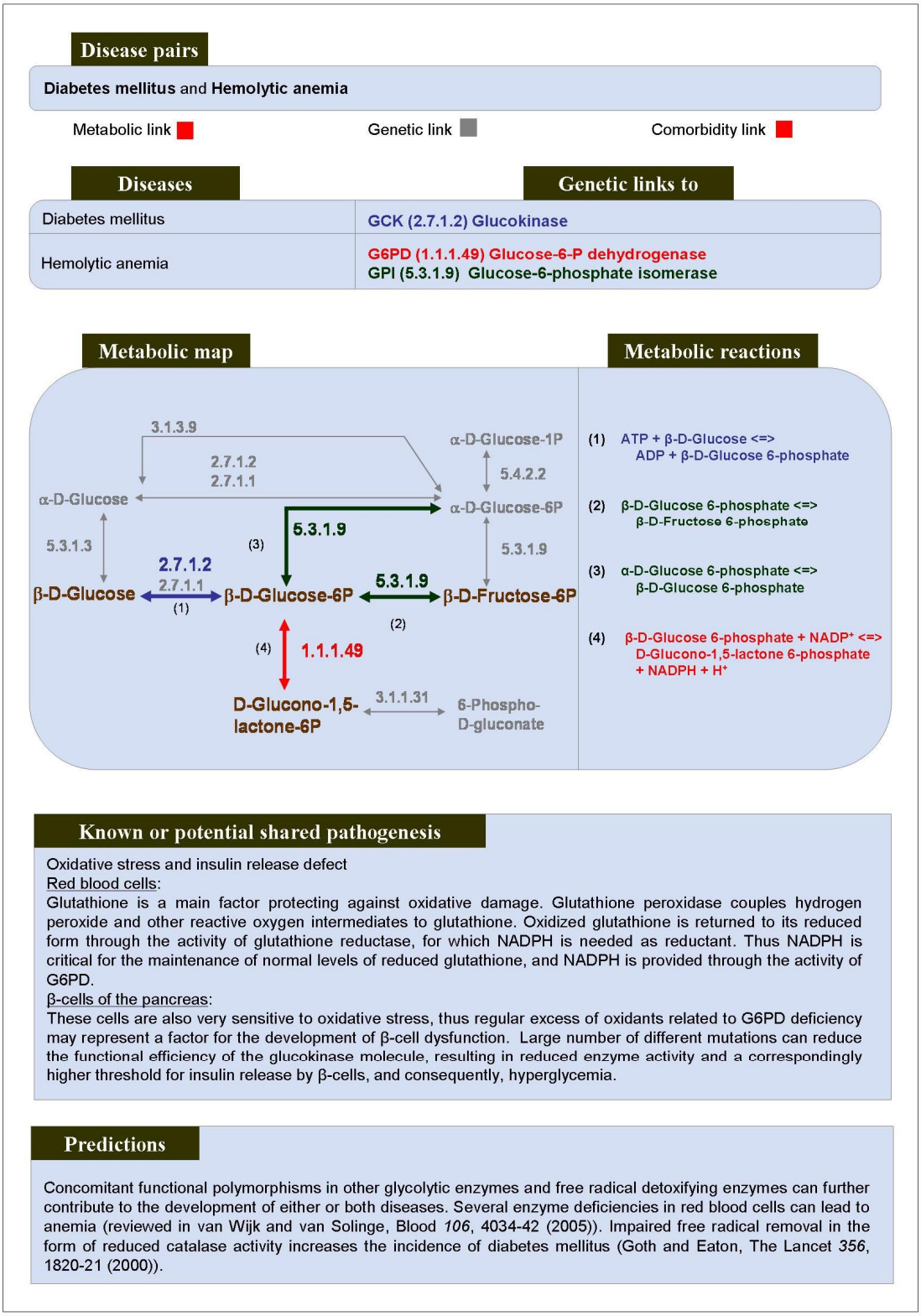


Fig. S7. Disease comorbidity report card for diabetes mellitus and hemolytic anemia. These two diseases, having metabolic links and no genetic link, have comorbidity 0.0038 ($P < 10^{-8}$). Shown in the report card are their associated genes (enzymes), the metabolic map in which their adjacent reactions are highlighted and colored differently, their known pathogenesis, and the implication of the metabolic links to the pathogenesis. Similar report cards for disease pairs are shown in Figs. S8–S11.

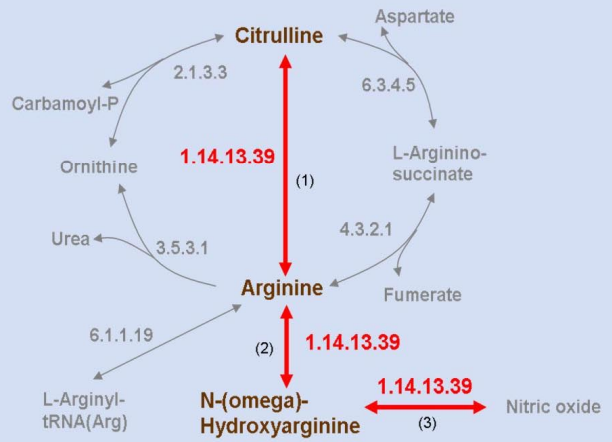
Disease pairs

Hypertension and **Coronary spasm**

Metabolic link ■ Genetic link ■ Comorbidity link ■

Diseases	Genetic links to
Hypertension	NOS3 (1.14.13.39) Nitric oxide synthase (endothelial)
Coronary spasm	

Metabolic map



Metabolic reactions

(1) L-Arginine + Oxygen + NADPH <=> Nitric oxide + L-Citrulline + $NADP^+$

(2) L-Arginine + Oxygen + NADPH <=> N-(omega)-Hydroxyarginine + $NADP^+$ + H2O

(3) NADPH + 2 N-(omega)-Hydroxyarginine + Oxygen + H^+ <=> $NADP^+$ + 2 Nitric oxide + 2 L-Citrulline

Known or potential shared pathogenesis

There are a large number of studies on the co-occurance of the two diseases. Some of the underlying pathophysiology, namely vasoconstriction, is shared between the two diseases. The tone of the smooth muscle layer under the endothelial layer is a key determinant of the tone of all blood vessels in the human body. Endothelial cells release several types of molecules that affect the underlying smooth muscle cell layer, of which the most important one is nitric oxide. In turn several types of protein complexes on the smooth muscle surface play a pivotal role on regulating the contractile state of the smooth muscle cells, including K^+ channels and Ca^{2+} channels. Nitric oxide, directly or indirectly can regulate the activity of both.

Predictions

Additional factors affecting vasoconstriction may contribute to both diseases' pathogenesis.

Fig. S8. Disease report card for hypertension and coronary spasm.

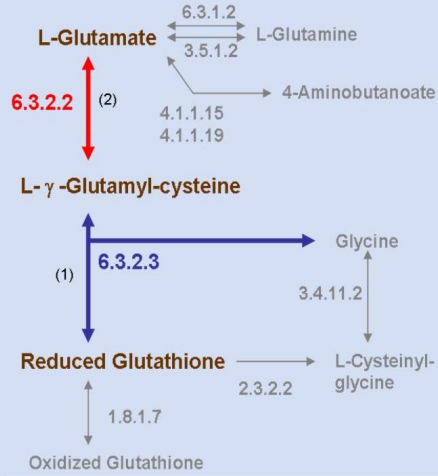
Disease pairs

Glutathione synthetase deficiency and **Myocardial infarction**

Metabolic link ■ Genetic link ■ Comorbidity link ■

Diseases	Genetic links to
Glutathione synthetase deficiency	GSS (6.3.2.3) Glutathione synthetase
Myocardial infarction	GCLM (6.3.2.2) Glutamate-cysteine ligase

Metabolic map



Metabolic reactions

(1) ATP + γ-L-Glutamyl-L-cysteine + Glycine <=> ADP + Orthophosphate + Glutathione

(2) ATP + L-Glutamate + L-Cysteine <=> ADP + Orthophosphate + γ-L-Glutamyl-L-cysteine

Known or potential shared pathogenesis

Currently there is no published evidence for the two diseases co-occuring.

The tone of the smooth muscle layer under the endothelial layer is a key determinant of the tone of all blood vessels, including the coronary arteries of the human heart. Endothelial cells release several types of molecules that affect the underlying smooth muscle cell layer, of which the most important one is nitric oxide (NO). The normal production of NO is in part dependent on the normal redox state of the endothelial cells, and the normal concentration of reduced glutathione is a significant determinant of this. Indeed, reduced glutathione levels (together with several other factors, such as elevated homocysteine) are consistently associated with reduced coronary blood flow, and polymorphism in the regulatory subunit of the glutamate-cysteine ligase (the rate limiting enzyme of glutathione synthesis) has been shown to be associated with impairment of NO-mediated coronary vasomotor function (Nakamura *et al*, Circulation *108*, 1425-7 (2003)). Such symptoms are also expected in glutathione synthase deficiency, but there are very few described patients with GSS deficiency, and most have very serious additional symptoms (reviewed in Njalsson, Cell. Mol. Life Sci. *62*, 1938-45 (2005)).

Predictions

Additional factors affecting the redox state of the cell may contribute to both diseases' pathogenesis.

Fig. S9. Disease report card for glutathione synthetase deficiency and myocardial infarction.

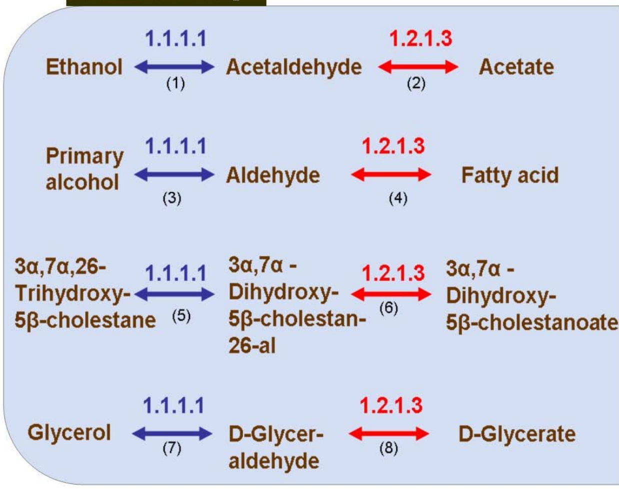
Disease pairs

Alcoholism and **Epilepsy**

Metabolic link ■ Genetic link ■ Comorbidity link ■

Diseases	Genetic links to
Alcoholism	ADH1B (1.1.1.1) Alcohol dehydrogenase
Epilepsy	ALDH7A1 (1.2.1.3) Aldehyde dehydrogenase

Metabolic map



Metabolic reactions

(1) Ethanol + NAD+ <=> Acetaldehyde + NADH + H^+

(2) Acetaldehyde + NAD(P)+ + H2O <=> Acetate + NAD(P)H + H^+

(3) Primary alcohol + NAD+ <=> Aldehyde + NADH + H^+

(4) Aldehyde + NAD+ + H2O <=> Fatty acid + NADH

(5) 3alpha,7alpha,26-Trihydroxy-5beta-cholestane + NAD+ <=> 3alpha,7alpha-Dihydroxy-5beta-cholestan-26-al + NADH

(6) 3alpha,7alpha-Dihydroxy-5beta-cholestan-26-al + NAD+ + H2O <=> 3alpha,7alpha-Dihydroxy-5beta-cholestanate + NADH + H^+

(7) Glycerol + NAD(P)+ <=> D-Glyceraldehyde + NAD(P)H + H^+

(8) D-Glyceraldehyde + NAD+ + H2O <=> D-Glycerate + NADH + H+

Known or potential shared pathogenesis

There is anecdotal evidence for the two diseases co-occuring, i.e. suggesting that prevalence of epilepsy among alcoholics is significantly higher than in the general population, and also that alcoholism may be more prevalent among epileptic patients than in the general population. However, there is no firm data for either.

Fast synaptic inhibition in the brain and spinal cord is mediated largely by ionotrophic –γ-aminobutyric acid (GABA) receptors, which is a group of ligand-gated ion channels. These receptors are involved in a number of central nervous system diseases, including epilepsy and alcoholism. In turn, there are several drugs that are direct modulators of GABA receptors.

In the normal brain, a neurotransmitter called glutamate is released when one experiences a desire for something. This 'go' signal is counteracted by GABA, and the ratio of glutamate to GABA signal determines the 'craving' response. If there is either an excess of glutamate signal or reduced GABA signal addiction can occur.

A drug called topiramate simultaneously reduces the effect of glutamate and enhances the effect of GABA, and is used for the treatment of epileptic seizures. Patients have noted that it also helps to fight their alcoholism, and the drug is currently use 'off label' in alcoholics.

Predictions

Additional factors affecting the levels of glutamate and GABA in the brain may contribute to the pathogenesis of both diseases.

Fig. S10. Disease report card for alcoholism and epilepsy.

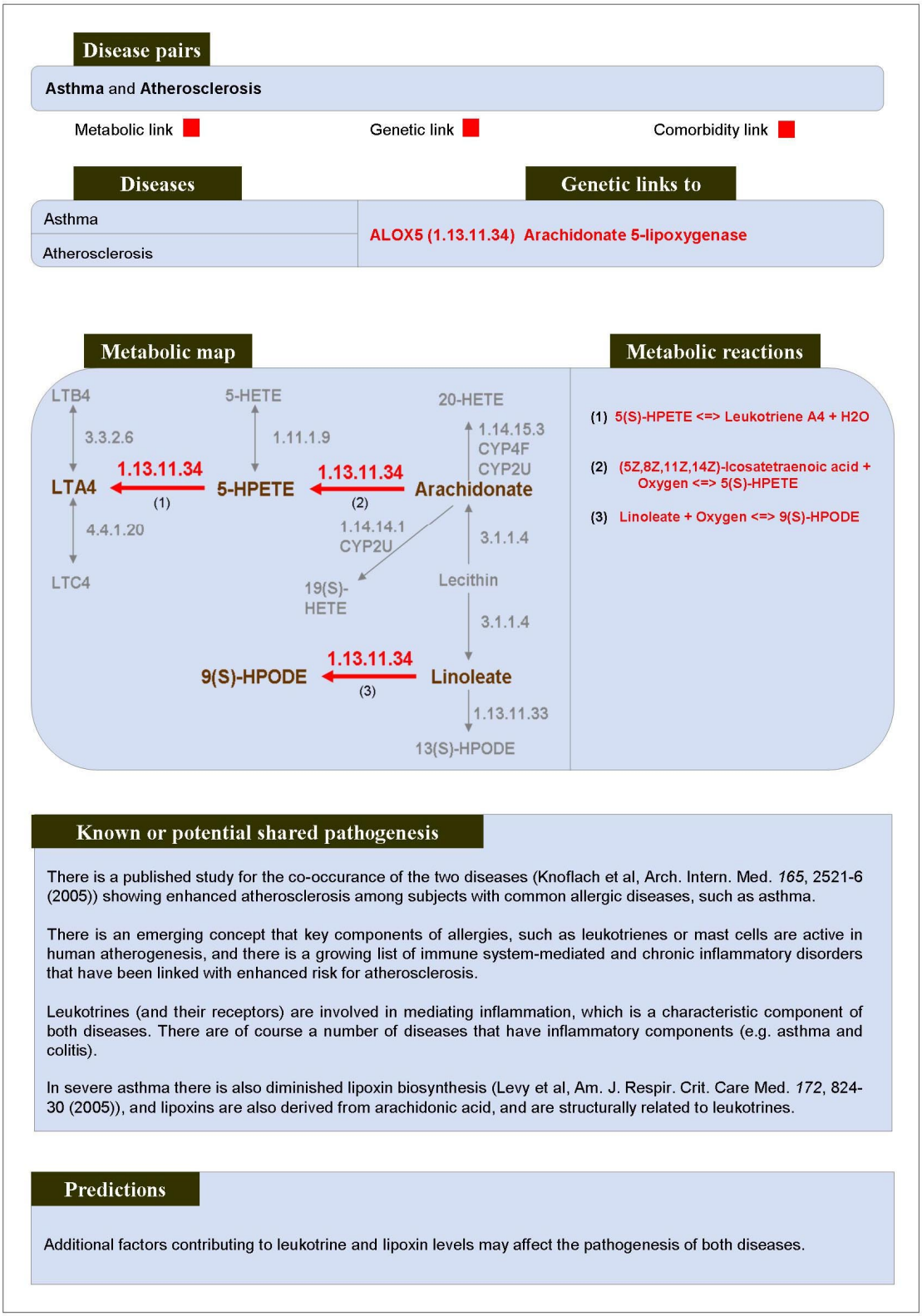


Fig. S11. Disease report card for asthma and atherosclerosis.

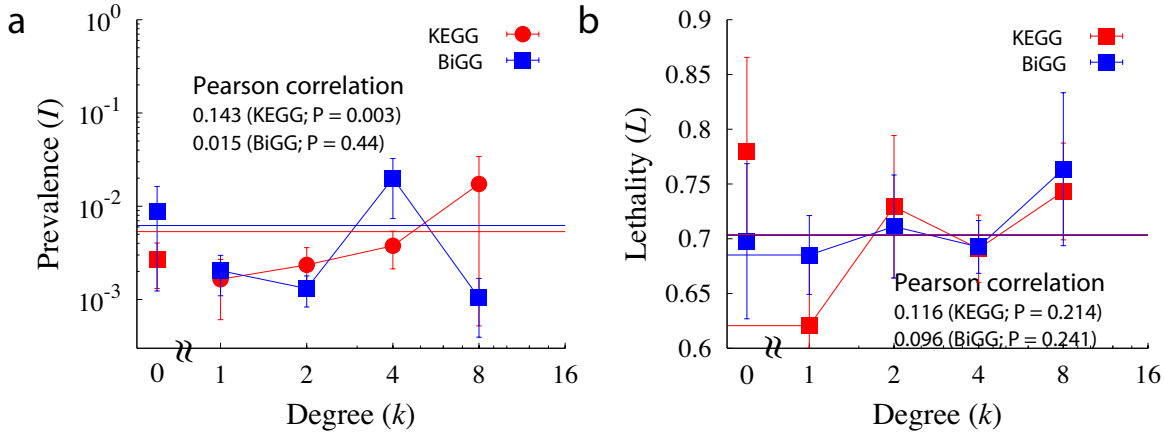


Fig. S12. Prevalence and mortality of monogenic diseases. Among 337 (378) metabolism-related diseases in the MDN according to the KEGG (BiGG) database, 118 (130) diseases are associated with the mutation of a single gene, respectively. The positive correlations between the degree in the MDN and the prevalence/mortality, as shown in Fig. 3 *c* and *e*, are identified also for these monogenic diseases. (*a*) Prevalence as a function of the degree of monogenic diseases in the MDN. The Pearson correlation coefficient is 0.143 for KEGG databases and 0.015 for BiGG with P values 0.003 and 0.44, respectively. (*b*) Mortality as a function of the degree of monogenic diseases in the MDN. The Pearson correlation coefficient is 0.116 for KEGG databases and 0.096 for BiGG with *P* values 0.21 and 0.24, respectively.

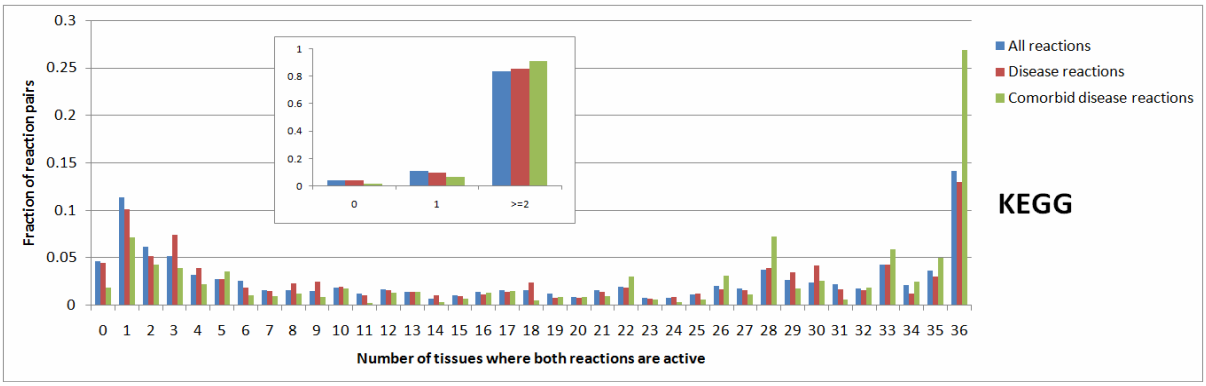


Fig. S13. Distribution of the number of tissues in which two metabolic reactions are active together in the human metabolic network from the KEGG database. Three different distributions are shown, each of which is obtained for all pairs of reactions (blue), pairs of disease reactions (red), and pairs of disease reactions whose associated diseases show high comorbidity (green).

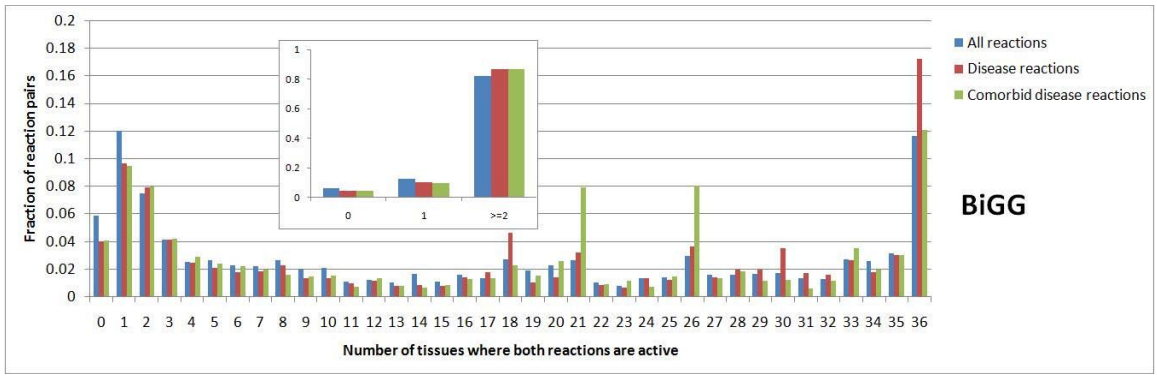


Fig. S14. Distribution of the number of tissues in which two metabolic reactions are active together in the human metabolic network from the BiGG database. Three different distributions are shown, each of which is obtained for all pairs of reactions (blue), pairs of disease reactions (red), bad pairs of disease reactions whose associated diseases show high comorbidity (green).

Table S1. Disease pairs that have the highest comorbidity and are connected in both KEGG and BiGG database

Disease1	Disease2	Coincidence (Expected value)	Comorbidity (Maximum possible comorbidity)
Diabetes mellitus	Obesity	115,638 (53,151)	8.3266×10^{-2} (0.35316)
Hypertension	Coronary spasms, susceptibility to	326,513 (225,637)	7.4141×10^{-2} (0.32667)
Hyperthyroidism, congenital	Total iodide organification defect	9,455 (1,849)	5.0372×10^{-2} (0.22563)
Endometrial carcinoma	Ovarian cancer	1,359 (129)	3.0114×10^{-2} (0.73863)
Glutathione synthetase deficiency	Myocardial infarcation, susceptibility to	4,900 (1,725)	2.1414×10^{-2} (0.95010)
Lhermitte-Duclos syndrome	Oligodendroglioma	109 (3)	1.6810×10^{-2} (0.80107)
Alcoholism, susceptibility to	Epilepsy	2,038 (656)	1.5058×10^{-2} (0.58752)
Goiter	Hyperthyroidism, congenital	426 (52)	1.4343×10^{-2} (0.72645)
Goiter	Total iodide organification defect	2,489 (977)	1.3767×10^{-2} (0.16397)
Diabetes mellitus	Hyperinsulinemic hypoglycemia, familial, 3	711 (175)	1.2320×10^{-2} (0.019991)
Enolase-β deficiency	Myopathy	107 (7)	1.0519×10^{-2} (0.37852)
Aldosteronism, glucocorticoid-remediable	Hypoaldosteronism, congenital	58 (3)	9.0245×10^{-3} (0.32134)
Favism	Hemolytic anemia	13 (0.2)	7.8440×10^{-3} (0.22424)
Asthma	Atopy	341 (90)	7.4047×10^{-3} (0.11757)
Aldosteronism, glucocorticoid-remediable	Low renin hypertension, susceptibility to	1,148 (662)	6.4234×10^{-3} (0.017235)
Asthma	Atherosclerosis, susceptibility to	7,084 (5,889)	4.4056×10^{-3} (0.96029)
Glutathione synthetase deficiency	Hemolytic anemia	210 (80)	4.0737×10^{-3} (0.22471)
Colon adenocarcinoma	Ovarian cancer	816 (505)	3.8768×10^{-3} (0.370902)
Diabetes mellitus	Hemolytic anemia	1,656 (1,215)	3.8373×10^{-3} (0.05276)
Colon adenocarcinoma	Cowden disease	93 (25)	3.8059×10^{-3} (0.082376)

This is the subset of Dataset S5, listing the diseases that are connected in both the MDN based on the KEGG and BiGG databases, and cooccur in >10 patients and have the highest comorbidity indices.

Other Supporting Information Files

Dataset S1
Dataset S2
Dataset S3
Dataset S4
Dataset S5