1 Predicting functional neuroanatomical maps from 2 fusing brain networks with genetic information

3 Florian Ganglberger[1], Joanna Kaczanowska[2], Josef M. Penninger[3], Andreas Hess[4], Katja 4 Bühler[1*] and Wulf Haubensak[2*]

5 [1] VRVis Research Center, Donau-City Strasse 1, 1220 Vienna, Austria

6 [2] Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Dr. Bohr-Gasse 7 7, 1030 Vienna, Austria

8 [3] Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna 9 Biocenter (VBC), 1030 Vienna, Austria

10 [4] Institute of Experimental and Clinical Pharmacology and Toxicology, Friedrich-Alexander 11 University Erlangen-Nuremberg, Fahrstrasse 17, 91054 Erlangen, Germany

12 [*]To whom correspondence should be addressed.
13

14 Summary

15 A central aim, from basic neuroscience to psychiatry, is to resolve how genes control 16 brain circuitry and behavior. This is experimentally hard, since most brain functions 17 and behaviors are controlled by multiple genes. In low throughput, one gene at a time, 18 experiments, it is therefore difficult to delineate the neural circuitry through which 19 these sets of genes express their behavioral effects. The increasing amount of publicly 20 available brain and genetic data offers a rich source that could be mined to address this 21 problem computationally. However, most computational approaches are not tailored to 22 reflect functional synergies in brain circuitry accumulating within sets of genes. Here, 23 we developed an algorithm that fuses gene expression and connectivity data with 24 functional genetic meta data and exploits such cumulative effects to predict 25 neuroanatomical maps for multigenic functions. These maps recapture known 26 functional anatomical annotations from literature and functional MRI data. When 27 applied to meta data from mouse QTLs and human neuropsychiatric databases, our 28 method predicts functional maps underlying behavioral or psychiatric traits. We show 29 that it is possible to predict functional neuroanatomy from mouse and human genetic 30 meta data and provide a discovery tool for high throughput functional exploration of 31 brain anatomy *in silico*.

32 Introduction

33 The wealth of data from brain initiatives and the increasing amount of functional genetic 34 information creates opportunities to mine these resources for insights into the genetic and 35 neuronal organization of brain function and behavior. Recent studies correlated brain gene 36 expression maps with structural information to enhance our understanding of genetic and 37 anatomical parcellations of the brain (1, 2) and its functional networks (3). These studies have 38 been used, for instance, to explore development and physiological regulation of structural 39 connectivity and extract functional networks *in silico* (Supplementary Note 1). Collectively,

40   these results suggest that functional genetic information, brain gene expression data and
41   connectomes can be successfully used for functional exploration of the brain (Supplementary
42   Fig. 1).

43   Here, we mine these resources to understand how genes control behavior. A major challenge
44   in this regard is that behaviors are inherently multigenic and, consequently, identifying the
45   neural networks through which these gene sets interact to express that function is not trivial.
46   Discovery tools that give computational predictions would provide an ideal entry point into
47   this problem.

48   Most established approaches that map genetic information to brain data relate gene co-
49   expression correlation of functionally grouped genes with structural connectivity (2–5).
50   Correlative analysis primarily dissects brain organization based on the similarities of regional
51   gene expression (Supplementary Note 1). It primarily reflects transcriptomic similarities,
52   globally or for subsets of genes, but it is not tailored to directly predict functional synergies
53   accumulating over multiple genes. Motivated by this methodological gap, we sought to
54   develop algorithms that fuse genetic information (sets of functionally related genes) with
55   brain data to generate functional neuroanatomical maps underlying a given brain function or
56   behavior *in silico*.
57
58   We hypothesize that functional synergies of gene sets are best reflected in their cumulative
59   weights on higher order features of structural (connectomes) or functional (resting state) brain
60   networks. Based on this, we developed a method that generates functional neuroanatomical
61   maps of functionally related gene sets from literature meta-analyses or genetic databases. We
62   demonstrate that cumulative gene expression reflects those functional synergies. Calculating
63   the effects of cumulative gene expression on different network measures (6, 7) proved to be
64   sufficient for predicting functional neuroanatomy of multigenic brain functions and behavior.
65   When applied to gene sets from genome wide association studies, quantitative trait loci
66   (QTL) analyses or neurogenetic databases, these calculations allowed to predict brain circuits
67   underlying complex behavioral traits in mice and human.

Results

69   The method was developed on the Allen Mouse Brain Atlas (AMBA) gene expression and
70   connectivity data framework (8, 9), a widely used mouse brain database. The mouse brain is
71   currently the most advanced template for integrated network studies of mammalian brains
72   with extensive gene expression and connectomic information available (8, 9). However, the
73   method as such is general and can be applied straight forward to data from any other species
74   such as human. The code has been optimized for low cost parallel computing.

75   Specifically, our method employs genetic-functional associations as inputs for weighting
76   brain data. We fused a set of genes associated with a given brain function or behavior with
77   gene expression maps and connectome (as structural brain network) (Fig. 1). We define the
78   input set **T** of genes out of a genome-wide set **G**. The spatial brain gene expression data is
79   imported pre-aligned to a common reference space from AMBA. The gene expression data
80   consists of ordered lists of gene expression densities (10) retrieved from the AMBA for a set
81   of spatial grid positions $\mathbf{D} = \{d_i\}_{i=1..n}$ and stored as gene expression density volumes $\mathbf{D(T)}$ and
82   $\mathbf{D(G)}$. Gene expression density is not location invariant. For example, cortical and thalamic
83   areas have a higher mean gene expression density than the rest of the brain. Spatial bias
84   introduced by this variance was compensated by the standardization (Z-Score) of

85   $\mathbf{D}(\mathbf{T})$ genome-wide, such that expression density distributions at every spatial position are
86   standard-normal distributed over $\mathbf{G}$. Subsequently, these data sets were standardized in their
87   spatial distribution pattern to adjust for differences between genes within the overall brain
88   expression density.

89   Next, we sought to determine the cumulative genetic weight of $\mathbf{T}$ in $\mathbf{D}$ and calculated the
90   synergy $\mathbf{S}$, defined as the trimmed mean of the normalized $\mathbf{D}$ for all genes in set $\mathbf{T}$. Trimming
91   reduced sampling artifacts in gene density maps, like image artifacts that appear as outliers
92   with high density scores (e.g. air bubbles) (11). The functional relation between genes and
93   neuroanatomy is expressed by weighting either incoming or outgoing connections of every
94   spatial sample point by $\mathbf{S}$. Given the directed AMBA connectome as a connectivity
95   matrix $\mathbf{C} \in \mathbf{R}^{n \times n}$ (where rows represent source regions, and columns target regions), an
96   incoming- or outgoing-weighted connectome is defined as the row- or column-wise
97   multiplication of $\mathbf{C}$ by $\mathbf{S}$. To account for higher order synergies within functional maps, we
98   computed those maps from incoming and outgoing node strengths as local network measures
99   (12) in the weighted connectomes. For statistical evaluation, we compared the position-
100 wise node strength measures to randomly drawn gene sets (n=1000) from the genome-wide
101 set $\mathbf{G}$ by t-tests (Fig. 1). We adjusted the False Discovery Rate (FDR) of the p-values with the
102 Benjamini-Hochberg (13) method. The results in this paper are all significant under a FDR
103 <5% (unless indicated otherwise). Ultimately, these operations generated a p-value map (a p-
104 value for every sampling position) for every effect and brain function. To add structural
105 context, these maps were combined (minimum p-value of effects) and projected onto the
106 connectome, building structural networks of functionally weighted nodes that are functionally
107 related to the input gene set. A detailed description can be found in the Supplementary
108 Experimental Procedures.

109 To assess if this computational approach allows to identify function-specific brain circuitry,
110 we focused on several well-studied gene sets, for which functional associations and
111 functional neuroanatomy are comprehensively documented: genes associated with
112 dopaminergic signaling, social behavior, feeding, hypothalamic–pituitary–adrenal (HPA)
113 stress axis and synaptic plasticity. With these gene sets, we recaptured known functional
114 neuroanatomy from literature.

115 For instance, genes associated with social behavior recapitulated their known functional
116 neuroanatomy (Fig. 2*A*, Supplementary Data 1) (14–20). Similarly, we were able to pick up
117 the functional neuroanatomy (Supplementary Data 3 Case 1-10*A,B,C*, Supplementary Data 1)
118 for other functionally-associated gene sets (Supplementary Data 3 Case 1-10*D*) including
119 dopamine (DA) signaling, which revealed the classical DA reward VTA-ACB pathway and
120 also motor-related connections like SN-GP (21–24). The method allowed detecting the
121 known feeding-related neuroanatomy based on genes associated with feeding, like orexin,
122 neuropeptide Y (NPY), Agouti related protein (AgRP), proopiomelanocortin (POMC),
123 melanocortin or leptin receptors (25–28). Different stress and fear/anxiety-related genes
124 accumulate in the HPA axis, areas involved in control and regulation of stress and brain
125 regions involved in processing fear/anxiety (29–34). We also investigated gene sets for
126 synaptic plasticity, learning and memory. As expected, these genes highlight major sites of
127 behavioral and functional plasticity in the brain (e.g., cortex, hippocampus, amygdala) (35–
128 44).

129 To assess these predictions quantitatively, we collected the ground truth in form of network
130 nodes representing regions functionally associated with these 10 gene sets from literature

131 (Supplementary Data 2). We calculated the F_1-score (45) of precision and recall for a binary
132 classification of the ordered voxel-wise p-values. We used this with first order network
133 measures (expression site; genetic weight at the node itself) and second order network
134 measures (incoming/outgoing node strength from/to nodes with accumulated genetic weight,
135 as well as Hub score, Authority score, Closeness, Betweenness, and Eigencentrality) (Fig.
136 2*B*). The computational predictions correlated significantly with the known functional
137 neuroanatomy from literature (Fig. 2*B*, *bottom*, right bar), indicating that our method
138 assembles meaningful functional neuroanatomical maps from genetic data.

139 The predictive power increased from first order measures (Fig. 2*B*, *bottom*, middle bar) to
140 second order measures (Fig. 2*B*, *bottom*, right bar). This indicates that second order network
141 measures detected regions not identified by gene expression alone, yet are integrated within
142 the same neuroanatomical map. Results for node strength showed that the prediction accuracy
143 was superior to other network measures, and is therefore sufficient for further analysis.
144 Importantly, our approach is calculated at 100 μm voxel resolution, free from *a priori*
145 constraints from anatomical annotations and fully compatible with small rodent MRI. Thus, it
146 is suitable to refine structure-function relationships beyond neuroanatomical scales and has
147 the potential to identify additional nodes and subdivisions within predefined anatomical
148 regions with possible distinct physiological functions.

149 To further support our findings, we overlayed computed functional maps with those obtained
150 experimentally with fMRI. Important in the context of this paper, pain data offers the
151 possibility to link genetics with actual fMRI (46–48) in mice. In fact, for the pain-related
152 gene sets (Supplementary Note 2, Supplementary Table 3 and Supplementary Data 3 Case
153 11-30d), the *in silico* predicted functional maps in mouse brain were reproducing large
154 portions of the functional neuroanatomy observed with Blood-Oxygen-Level-Dependent
155 functional magnetic resonance imaging (BOLD fMRI data, warped onto the AMBA reference
156 space by optimized ANTS (49) parametrization) *in vivo* (Fig. 3*A*, b). This further
157 substantiates the validity of our approach. While our method seemed to fit best with sets of
158 >4 genes (Supplementary Fig. 2), predictions were also informative at the single-gene level.
159 Functional imaging data of Cacna2d3 mutants, a highly conserved pain gene, revealed altered
160 thalamo-cortical connectivity and synesthesia after thermal stimulation in mutant mice (50).
161 The predicted maps computed from Cacna2d3 alone (Fig. 3*A*, *top right*) recaptured pain
162 functional neuroanatomy from fMRI (Fig. 3*A*, *bottom left*, 3*B*) and pain maps that are
163 affected by this gene (Fig. 3*A*, *bottom right*, Fig. 3*B*). Nevertheless, the single gene
164 operations will depend heavily on the gene itself, and so we recommend to use gene sets for
165 the most efficient and accurate functional neuroanatomy integration.

166 Based on these results, we explored yet unknown or only partially described effector
167 networks of behavioral traits investigated in genetic screens or association studies. One of the
168 challenges is that behavioral traits are largely multigenic and identifying the neural circuitry
169 through which these traits are expressed is difficult. We expanded our analysis on pain and
170 included fear/anxiety and autism spectrum disorder (ASD) gene sets (Supplementary Note 2)
171 from publically available databases and published meta-studies (Supplementary Table 3). In
172 some cases, large gene sets were clustered using the DAVID platform to parcellate them into
173 functional category-linked subsets, and so in those cases genes are not only related by the
174 analyzed trait, but also regarding sub-functions annotated in the database. When supplied
175 with these gene sets, our computational method extracted meaningful functional maps
176 (Supplementary Data 3 Case 11-30). These maps, of which node-wise comparisons are in line
177 with their functional annotation from literature, give a comprehensive representation of

178 functional genetic synergies underlying the respective trait (Fig. 4*A*, green squares).
179 Interestingly, we also identified nodes so far not clearly linked to investigated functions,
180 therefore extracting potential novel functional elements (Fig. 4*A*, blue squares). These nodes
181 might be part of the same functional network and participate in shaping the internal states of
182 the mammalian brain.

183 Extending our approach to human template based on resting state networks from fMRI (as
184 functional brain networks) demonstrated that the methodology can be generalized to other
185 species. Cross-validation with the meta-studies (Supplementary Data 4, Supplementary Table
186 2) reveals similar findings for both (Fig. 4*A*,b), demonstrating its versatility for functional
187 exploration of the human brain in health and disease *in silico*.

Discussion

189 We have developed a computational method to integrate genetic, gene expression and
190 connectomic information from brain and genomic initiatives for rapid functional exploration
191 of the brain *in silico*. We found that, in the brain, functionally related genes are not
192 distributed at random but assemble into specific maps, which recapitulate functional
193 anatomical annotations or functional data from fMRI. Cumulative effects, from expression
194 sites alone (Fig. 2*B*, red bar), reflect functional synergies within functionally related genes,
195 which are not directly fitted by transcriptomic similarities, usually derived from correlative
196 analysis (Supplementary Note 1). The predictions further improved by second order network
197 measures, which incorporate functional synergies of local gene expression that manifest in
198 the context of higher-order interactions within the brain architecture. Incoming/Outgoing
199 node strength (Fig. 2*B*, green bar) performed best, but not significantly better than Hubs &
200 Authorities or Eigencentrality. This implies that nodes with the strongest effect on the
201 network are either primary expression sites, or source/target sites thereof. Betweenness and
202 Closeness, indicators of shortest paths in networks, outlined small distinctive nodes, that are
203 part of functional neuroanatomy, but failed to predict the entirety of functional
204 neuroanatomical annotations (explaining the seemingly random F_1-score in Fig. 2*B*). The
205 ground truth in its entirety might naturally be best explained by node strength, which reflects
206 compounded functional synergies of regions and their afferent and efferent connections.
207 Taken together, by fusing cumulative gene expression and best-fit network measures, we
208 provide an optimized tool that reliably predicts functional neuroanatomical maps from
209 genetic information.

210 When applied to gene sets from behavioral genetics, we demonstrated that our workflow can
211 extract putative effector network nodes as functional brain maps which can be used to explore
212 trait-specific circuitries. These explorations allowed to refine known functional
213 neuroanatomy (Fig. 4, green squares). For instance, the anatomy of thalamo-cortical
214 connections in pain processing can be dissected to fine anatomical resolution (e.g.,
215 Supplementary Data 3 Case 11*E*, red arrows, note layer specificity) which could not be
216 achieved with fMRI (Fig. 3*A*, wt). The method extracted a specific and strong connection
217 between PVT and central amygdala (Supplementary Data 3 Case 21*E*, red arrows).
218 Interestingly this connection recently emerged as central element in fear control (51, 52).
219 Similarly, for ASD, we identified many cortico-cortical connections (Supplementary Data 3
220 Case 24*E*,26-29*E*, red arrows) with prediction accuracy reaching individual layers. Among
221 similar lines, the method uncovered circuitry within regions functionally not yet associated
222 with specific traits (Fig. 4, blue squares). For instance, the functional association of visual
223 cortex with pain processing (53), motor cortex with startle response (54) and hypothalamic

224 circuitry in autism (55), whose roles are understudied in the context of the respective trait or
225 psychiatric condition, specifically at the fine anatomical or circuit level.

226 This can be particularly useful to pursue studies of causative role of genetic variance linked
227 to mental diseases with unknown ethiopathology or complex course/symptomatology (with
228 e.g., gene associations in GWAS studies as input). The method provides a holistic description
229 of the functional neuroanatomy of a given gene set related to a meta study or behavioral trait.
230 As such, it allows to rank order the most promising candidate regions. It has the potential to
231 refine the functional parcellation of the brain beyond anatomical resolution, especially when
232 performed with multiple functionally grouped gene sets at large scales. Importantly, the
233 candidate nodes, in particular those previously not associated with those conditions, can serve
234 as promising entry points for functional circuit dissection, e.g., with opto- and
235 pharmacogenetic methods.

236 The functional relation underlying our study can be exploited to associate gene sets with
237 specific brain functions or brain functions with specified gene sets (Supplementary Fig. 1).
238 Importantly, our strategy applies to other neural systems (beyond mouse and human) for
239 which genetic information, gene expression maps and connectomes are, or will be, available
240 and allows exploration of functional brain organization in cases where actual functional data
241 is difficult, if not impossible, to obtain.

Acknowledgments

Author Contributions

248 F.G., J.K. and W.H. conceived the method. F.G implemented the method, performed data
249 analysis and the quantitative validation. J.K performed the qualitative validation. J.P. and
250 A.H. provided fMRI data. F.G., J.K., A.H., K. B. and W.H. wrote the manuscript. K.B. and
251 W.H. jointly supervised the project.

Competing Financial Interests

253 The authors declare no competing financial interests.

254 **References**

255 1. French L, Pavlidis P (2011) Relationships between gene expression and brain wiring in
256   the adult rodent brain. *PLoS Comput Biol* 7(1).
257 2. French L, Tan PP, Pavlidis P (2011) Large-Scale Analysis of Gene Expression and
258   Connectivity in the Rodent Brain: Insights through Data Integration. *Front*
259   *Neuroinform* 5:12.
260 3. Richiardi J, Altmann A (2015) Correlated gene expression supports synchronous
261   activity in brain networks. *Science (80-)* 348(6240):11–14.
262 4. Rubinov M, Ypma RJF, Watson C, Bullmore ET (2015) Wiring cost and topological
263   participation of the mouse brain connectome. *Proc Natl Acad Sci* 112(32):201420315.
264 5. Whitaker KJ, et al. (2016) Adolescence is associated with genomically patterned
265   consolidation of the hubs of the human brain connectome. *Proc Natl Acad Sci* 113 (32
266   ):9105–9110.
267 6. Liska A, Galbusera A, Schwarz AJ, Gozzi A (2014) Functional connectivity hubs of
268   the mouse brain. *Neuroimage*.
269 7. Meunier D, Lambiotte R, Bullmore ET (2010) Modular and hierarchically modular
270   organization of brain networks. *Front Neurosci* 4(DEC).
271 8. Lein ES, et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain.
272   *Nature* 445(7124):168–176.
273 9. Oh SW, et al. (2014) A mesoscale connectome of the mouse brain. *Nature*
274   508(7495):207–214.
275 10. Lee C-K, et al. (2008) Quantitative methods for genome-scale analysis of in situ
276   hybridization and correlation with microarray data. *Genome Biol* 9(1):R23.
277 11. Bindhu P, Krishnapillai R, Thomas P, Jayanthi P (2013) Facts in artifacts. *J Oral*
278   *Maxillofac Pathol* 17(3):397–401.
279 12. Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: Uses
280   and interpretations. *Neuroimage* 52(3):1059–1069.
281 13. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical
282   and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* 57(1):289 – 300.
283 14. Kim Y, et al. (2015) Mapping social behavior-induced brain activation at cellular
284   resolution in the mouse. *Cell Rep* 10(2):292–305.
285 15. O'Connell LA, Hofmann HA (2011) The Vertebrate mesolimbic reward system and
286   social behavior network: A comparative synthesis. *J Comp Neurol* 519(18):3599–
287   3639.
288 16. Young LJ, Wang Z (2004) The neurobiology of pair bonding. *Nat Neurosci*
289   7(10):1048–1054.
290 17. Young LJ, Young AZM, Hammock EAD (2005) Anatomy and neurochemistry of the
291   pair bond. *Journal of Comparative Neurology*, pp 51–57.
292 18. Leshan RL, Pfaff DW (2014) The hypothalamic ventral premammillary nucleus: A key
293   site in leptin's regulation of reproduction. *J Chem Neuroanat* 61:239–247.
294 19. Marlin BJ xytoci. enables maternal behaviour by balancing cortical inhibition, Mitre
295   M, D'amour JA, Chao M V., Froemke RC (2015) Oxytocin Enables Maternal
296   Behavior by Balancing Cortical Inhibition. *Nature* 520(7548):499–504.
297 20. O'Connell L a., Hofmann H a. (2012) Evolution of a Vertebrate Social Decision-
298   Making Network. *Science (80-)* 336(6085):1154–1157.
299 21. Russo SJ, Nestler EJ (2013) The brain reward circuitry in mood disorders. *Nat Rev*
300   *Neurosci* 14(9):609–625.
301 22. Lammel S, Tye KM, Warden MR (2014) Progress in understanding mood disorders:
302   Optogenetic dissection of neural circuits. *Genes, Brain Behav* 13(1):38–51.

303 23. Bjoerklund A, Dunnett SB (2007) Dopamine neuron systems in the brain: an update.
304   *Trends Neurosci* 30(5):194–202.
305 24. Berridge KC, Kringelbach ML (2015) Pleasure Systems in the Brain. *Neuron*
306   86(3):646–664.
307 25. Betley JN, Cao ZFH, Ritola KD, Sternson SM (2013) Parallel, redundant circuit
308   organization for homeostatic control of feeding behavior. *Cell* 155(6):1337–1350.
309 26. Jennings JH, Rizzi G, Stamatakis AM, Ung RL, Stuber GD (2013) The inhibitory
310   circuit architecture of the lateral hypothalamus orchestrates feeding. *Science*
311   341(6153):1517–21.
312 27. Hardaway JA, Crowley NA, Bulik CM, Kash TL (2015) Integrated circuits and
313   molecular components for stress and feeding: Implications for eating disorders. *Genes,*
314   *Brain Behav* 14(1):85–97.
315 28. Wu Q, Clark MS, Palmiter RD (2012) Deciphering a neuronal circuit that mediates
316   appetite. *Nature* 483(7391):594–597.
317 29. Stoppel C, Albrecht A, Pape HC, Stork O (2006) Genes and neurons: Molecular
318   insights to fear and anxiety. *Genes, Brain Behav* 5(SUPPL. 2):34–47.
319 30. Tovote P, Fadok JP, Lüthi A (2015) Neuronal circuits for fear and anxiety. *Nat Rev*
320   *Neurosci* 16(6):317–331.
321 31. Herman JP, Cullinan WE (1997) Neurocircuitry of stress: Central control of the
322   hypothalamo-pituitary-adrenocortical axis. *Trends Neurosci* 20(2):78–84.
323 32. Smith SM, Vale WW (2006) The role of the hypothalamic-pituitary-adrenal axis in
324   neuroendocrine responses to stress. *Dialogues Clin Neurosci* 8(4):383–395.
325 33. Carhuatanta K a K, Shea CJ a, Herman JP, Jankord R (2014) Unique genetic loci
326   identified for emotional behavior in control and chronic stress conditions. *Front Behav*
327   *Neurosci* 8:341.
328 34. Steimer T (2002) The biology of fear- and anxiety-related behaviors. *Dialogues Clin*
329   *Neurosci* 4(3):231–249.
330 35. Ressler KJ, Paschall G, Zhou X, Davis M (2002) Regulation of synaptic plasticity
331   genes during consolidation of fear conditioning. *J Neurosci* 22(18):7892–7902.
332 36. Mineur YS, Crusio WE, Sluyter F (2004) Genetic dissection of learning and memory
333   in mice. *Neural Plast* 11(3-4):217–240.
334 37. Toyoda H, et al. (2011) Interplay of amygdala and cingulate plasticity in emotional
335   fear. *Neural Plast* 2011.
336 38. Pisani A, Centonze D, Bernardi G, Calabresi P (2005) Striatal synaptic plasticity:
337   Implications for motor learning and Parkinson's disease. *Mov Disord* 20(4):395–402.
338 39. Lee Y-S (2014) Genes and signaling pathways involved in memory enhancement in
339   mutant mice. *Mol Brain* 7(1):43.
340 40. Hasan MT, et al. (2013) Role of motor cortex NMDA receptors in learning-dependent
341   synaptic plasticity of behaving mice. *Nat Commun* 4:2258.
342 41. Kirkwood A, Bear MF (1995) Elementary forms of synaptic plasticity in the visual
343   cortex. *Biol Res* 28:73–80.
344 42. Todd TP, Bucci DJ (2015) Retrosplenial Cortex and Long-Term Memory: Molecules
345   to Behavior. *Neural Plast* 2015.
346 43. Castro-Alamancos M a, Donoghue JP, Connors BW (1995) Different forms of
347   synaptic plasticity in somatosensory and motor areas of the neocortex. *J Neurosci*
348   15(July):5324–5333.
349 44. Iriki A, Pavlides C, Keller A, Asanuma H (1989) Long-Term Potentiation in the Motor
350   Cortex. *Science (80-)* 245(4924):1385–7.
351 45. Van Rijsbergen CJ (1979) *Information Retrieval, 2nd edition*.
352 46. Hess A, Sergejeva M, Budinsky L, Zeilhofer HU, Brune K (2007) Imaging of

353   hyperalgesia in rats by functional MRI. *Eur J Pain* 11(1):109–119.

354   47.   Hess A, et al. (2011) Blockade of TNF-alpha rapidly inhibits pain responses in the
355   central nervous system. *Proc Natl Acad Sci U S A* 108(9):3731–3736.

356   48.   Heindl-Erdmann C, et al. (2010) Combining functional magnetic resonance imaging
357   with mouse genomics: new options in pain research. *Neuroreport* 21(1):29–33.

358   49.   Avants BB, Epstein CL, Grossman M, Gee JC (2008) Symmetric Diffeomorphic
359   Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly
360   and Neurodegenerative Brain. *Med Image Anal* 12(1):26–41.

361   50.   Neely GG, et al. (2010) A Genome-wide Drosophila screen for heat nociception
362   identifies α2δ3 as an evolutionarily conserved pain gene. *Cell* 143(4):628–638.

363   51.   Do-Monte FH, Quinones-Laracuente K, Quirk GJ (2015) A temporal shift in the
364   circuits mediating retrieval of fear memory. *Nature* 519(7544):460–463.

365   52.   Penzo MA, et al. (2015) The paraventricular thalamus controls a central amygdala fear
366   circuit. *Nature* 519(7544):455–459.

367   53.   Gopalakrishnan R, Burgess RC, Plow EB, Floden DP, Machado AG (2015) A
368   magnetoencephalography study of multi-modal processing of pain anticipation in
369   primary sensory cortices. *Neuroscience* 304:176–189.

370   54.   Kühn AA, Sharott A, Trottenberg T, Kupsch A, Brown P (2004) Motor cortex
371   inhibition induced by acoustic stimulation. *Exp Brain Res* 158(1):120–124.

372   55.   Kurth F, et al. (2011) Diminished Gray Matter Within the Hypothalamus in Autism
373   Disorder: A Potential Link to Hormonal Effects? *Biol Psychiatry* 70(3):278–282.

374

Ganglberger et al.
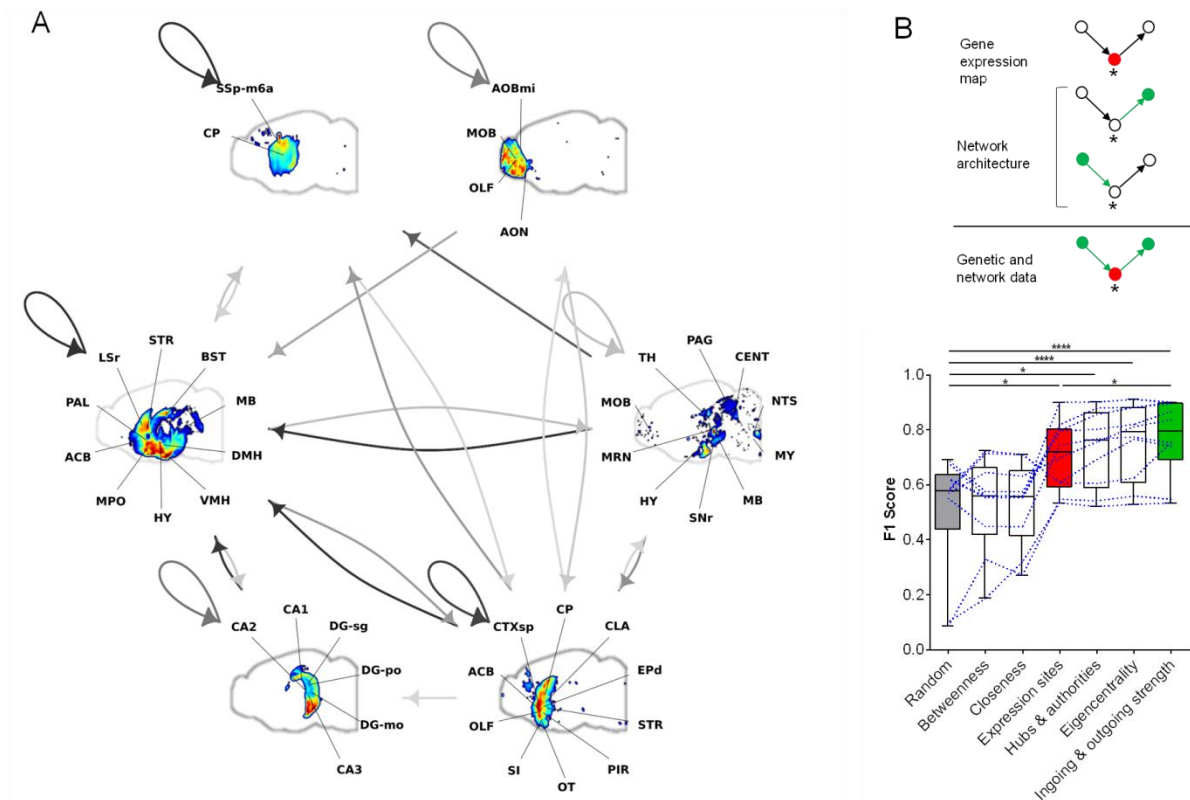Functional neuroanatomy from genes
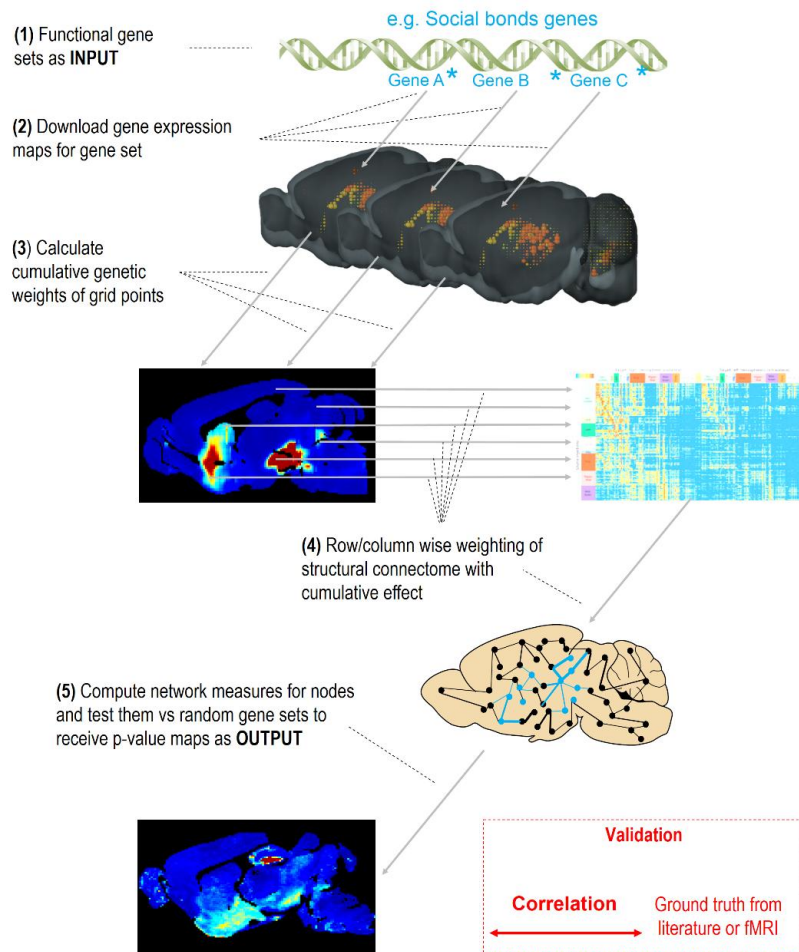
375

Figure 1. Computational workflow. A functionally-related gene set serves as input (1). For this gene set, gene expression data is retrieved (2), normalized and used to calculate a cumulative genetic effect (3). The cumulative effect is used to weight a structural connectivity matrix column or row wise (4). On the weighted network, network measures are computed and statistically evaluated by t-tests against a null distribution (network measures based on random gene sets) (5). The output is a voxel-wise p-value map for every network measure. The results can be evaluated by computing correlation with ground truth from literature or fMRI.

Figure 2. Recovery of known functional anatomy from test gene sets. (*A*) Clustered nodes of a functional anatomical map associated with a gene set for social behavior, overlayed with structural connectivity (grey arrows). The top-ranked networks include olfactory bulb (MOB), olfactory tubercle (OT), endopiriform nucleus (EPd), substantia innominata (SI), hypothalamus (HT) and hypothalamic nuclei (dorsomedial nucleus of the hypothalamus (DMH), lateral hypothalamic area (LHA), ventromedial hypothalamic nucleus (VMH)), hippocampus (particularly CA2 region), midbrain (MB), including periaqueductal gray (PAG), ventral tegmental area (VTA) and nucleus accumbens (ACB). The pseudo-color scale of the nodes (colored voxels) indicates the voxel-wise accumulation of genetic weights, the intensity of the edges (arrows) the structural strength of the connection between the nodes. Loops indicate within node connections. For a complete list of abbreviations see Supplementary Tab. 1. (*B*) *Top*, Integration of first and second order network measures. The asterisk indicates a node with accumulated genetic weight. Red and green indicate sites with increased weight in first and second order measures, respectively. *Bottom*, Node-wise comparison of predicted maps to ground truth for 10 test sets. F_1-scores increase from random classification to expression sites and to second order network measures significantly (Benjamini & Hochberg corrected One-way ANOVA on ranks, Ingoing & outgoing network strength vs Expression sites; $p<0.05$, Ingoing & outgoing network strength vs Random; $p<0.0001$, Expression sites vs Random; $p<0.05$, Eigencentrality vs Random; $p<0.0001$, Hubs & authorities vs Random; $p<0.05$). The individual F_1 scores for each prediction are shown as dotted lines. Bars indicate median and interquartile range. Incoming/Outgoing node strength, Hubs, Authorities, Closeness, Betweenness and Eigencentrality were tested, node strength showed the highest F_1 score.
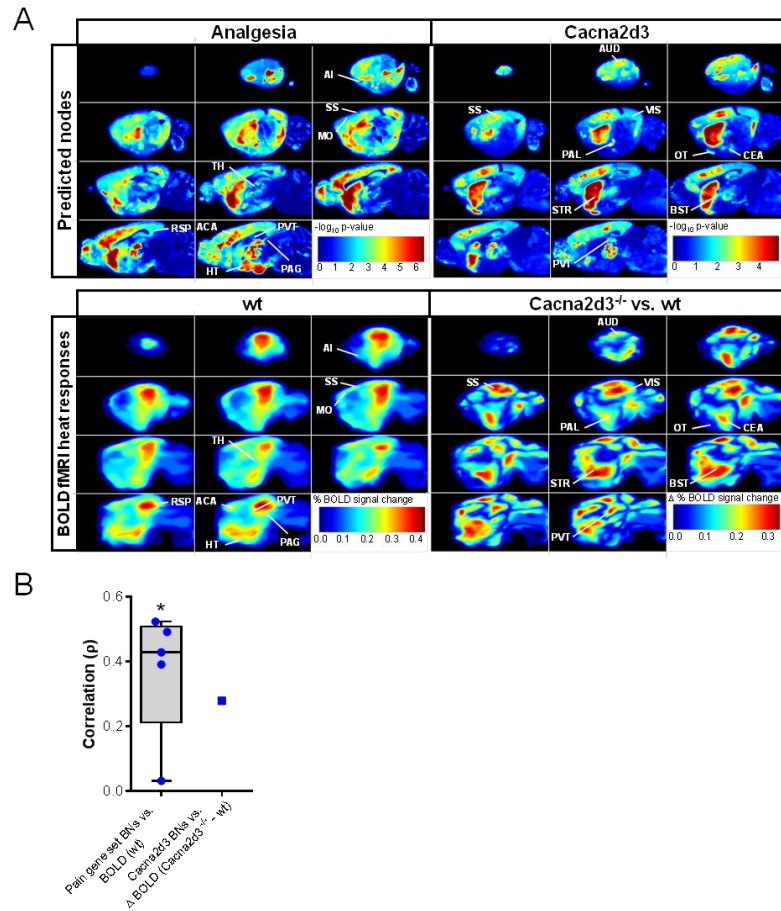
Figure 3. Computed functional maps correlate with BOLD fMRI of pain-related states. (*A*) Similarity of functional maps nodes predicted for analgesia gene sets and Cacna2d3 gene (*top*) to nodes with heat evoked fMRI responses (*bottom*). The highest ranked nodes include striatum (STR), paraventricular nucleus of thalamus (PVT), bed nuclei of stria terminalis (BST), pallidum (PAL), central amygdalar nucleus (CEA), sensory cortices (somatosensory areas (SS), visual areas (VIS), auditory areas (AUD)) and olfactory tubercle (OT) and correspond to those identified by fMRI. Color bars indicate $-\log_{10}$-scaled p-values (*top*) and heat stimulus responses (% BOLD signal changes) in wt animals (*bottom left*) or differences (Δ) in heat responses between Cacna2d3$^{-/-}$ and wt animals (% BOLD signal changes in Cacna2d3$^{-/-}$ - % BOLD signal changes in wt animals) (*bottom right*). For a detailed list of brain regions see Supplementary Tab. 1. (*B*) Voxel-wise Spearman correlations of p-value maps predicted from pain gene sets with BOLD fMRI responses. Bars indicate median and interquartile range of Spearman's ρ. Wilcoxon signed rank test against $\rho=0$ (n=5, W^+ (15) =15, W (15) =0, *p$_{one-tailed}$ <0.05).
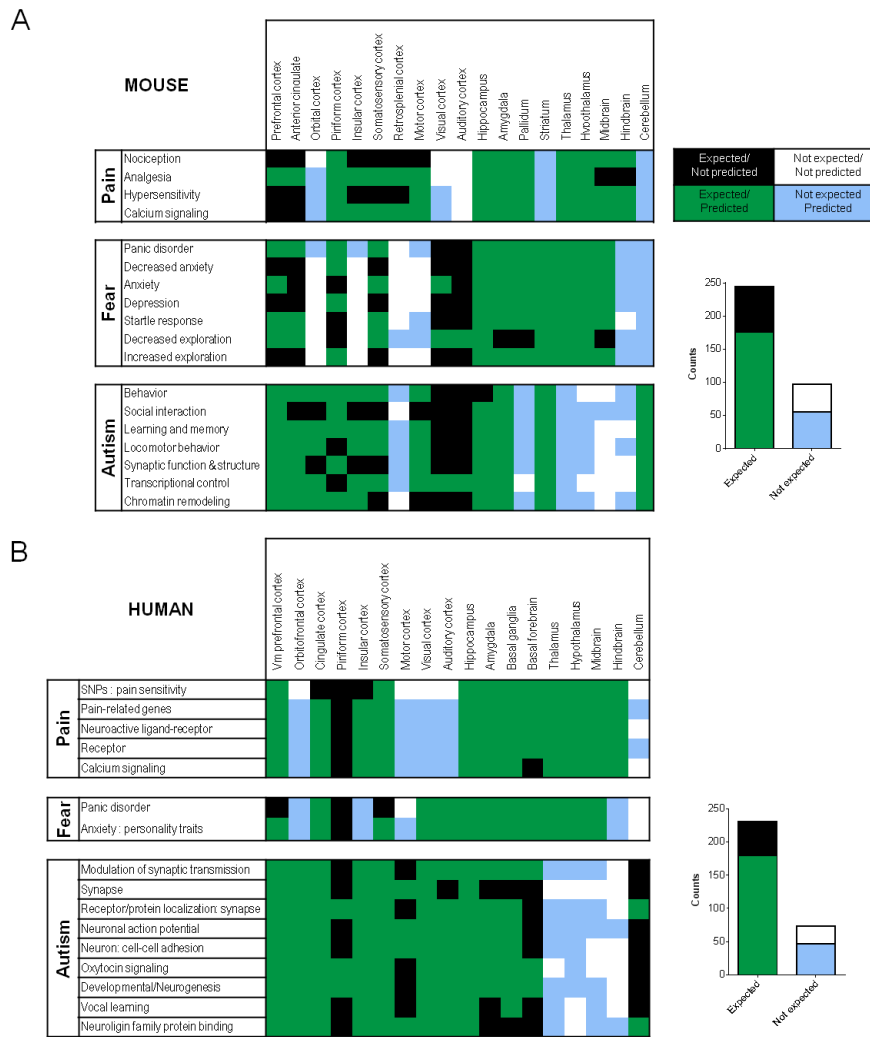
Figure 4. Predicting effector functional maps of behavioral traits from mouse and human genetic meta data. (*A*) *Left*, Node-wise comparison of predicted mouse functional anatomy for pain, fear and autism, divided into different functional subcategories, to functional neuroanatomical annotations from literature for the top 100 p-value ranked nodes. *Right*, Quantification of the qualitative assessment. There is a significant overlap between predicted maps and functional neuroanatomical annotation (n=342; Fisher's exact test, p=0.0121) (*B*) *Left,* Node-wise comparison of predicted human functional anatomy for pain, fear and autism, divided into different functional subcategories, to functional neuroanatomical annotations from literature for the top 100 p-value ranked nodes. *Right*, Quantification of the qualitative assessment. There is a significant overlap between predicted maps and functional neuroanatomical annotation (n=288; Fisher's exact test, p=0.0092).

Supplementary Information

Supplementary Figures 1-2.

Supplementary Data 1. P-values of first and second order effects for all cases based on region (mouse and human).

Supplementary Data 2. Ground truth generated from literature.

Supplementary Data 3. Functional neuroanatomical maps, significant regions and network visualization of all cases used in this paper for mouse.

Supplementary Data 4. Significant regions of all cases used in this paper for human.

Supplementary Table 1. Anatomical abbreviations.

Supplemental Experimental Procedures

Mouse Data

The mouse connectome was retrieved as (structural) connectivity from all 2173 available injection sites (state March 2016) to their target sites given as image data, detailing projections labeled by rAAV tracers via serial two-photon tomography (9). Those sites are added up to a connectivity matrix which covers about 15 percent of the right hemisphere as source regions, and about 100% as target regions. The AMBA connectome (right hemisphere injections) was mirrored onto (left hemisphere) AMBA gene expression data. In order to also take weak connections into account, the connectome was binarized by a threshold according to Oh, S. W. *et al.* (9), Extended Data Figure 7, that minimizes the amount of false positive connections. The gene expression density is interpolated to a 100 micron resolution to match the resolution of the connectome. A Matlab script for downloading the gene expression for **T** and for **G**, as well as the AMBA connectome is provided on request.

Human data

Gene expression by region retrieved from the Allen Human Brain Atlas (56). The Allen Institute provides an affine transformation to MNI152 (57) space by its API. We used resting state functional connectivity from the Human Connectome Project (58), which is also in MNI152space (57).

Mathematical description

Input data is a functionally related gene set, more precisely a certain brain function or behavioral trait represented as a list of genes. Spatial gene expression and connectomic data were retrieved from AMBA.

Data retrieval was performed via the AMBA API. It allows the download of 3D spatial gene

478 expression patterns(8) for available genes at given grid positions with a resolution of 200
479 microns.

480 We retrieve for n grid positions $\mathbf{P} = \{\mathbf{p}_i\}_{i=1,\ldots,n}$, $\mathbf{p}_i \in \mathbb{R}^3$ and each available gene g in the mouse
481 genome $\mathbf{G} = \{g_j\}_{j=1..m}$ (or at least a random drawn subset) the gene expression density

482 $\quad\quad d(\mathbf{p}_i, \mathbf{G}) := d_i(\mathbf{G}) = (d_{i1}, \ldots, d_{im}) \quad | \quad i=1,\ldots,n$

483 and store it as gene expression density volume

484 $\quad\quad \mathbf{D}(\mathbf{G}) = (d_1(\mathbf{G}), \ldots, d_n(\mathbf{G}))^T_{i=1..n} = (d_{ij})_{i=1..n\ j=1..m} \in \mathbb{R}^{n \times m}$

485 This is also done for the gene function/trait associated gene set $\mathbf{T} = \{t_k\}_{k=1,\ldots,l}$ being a subset
486 of $\mathbf{G}$, resulting in the expression density volume $\mathbf{D}(\mathbf{T}) \in \mathbb{R}^{n \times k}$.

487
488 <u>Normalization</u> of the function/trait specific expression density volume $\mathbf{D}(\mathbf{T})$ is performed
489 over the genomic as well as over the spatial domain. At first, standardization in the genome
490 space is performed so that every spatial sample point has a distribution of gene expression
491 densities with a mean of 0 and a standard deviation of 1 over the whole genome $\mathbf{G}$

492 $\quad\quad d_{ik}^{\text{gene normalized}} = (d_{ik} - \mu_i) / \sigma_i \quad\quad | \quad \forall\ d_{ik} \in \mathbf{D}(\mathbf{T})$

493 where $\mu_i = \mu((d_{ij})_{j=1,..,m})$ and $\sigma_i = \sigma(d_i((d_{ij})_{j=1,..,m})$, $d_{ij} \in \mathbf{D}(\mathbf{G}).$

494 This normalization compensated for spatial bias in the mean **density**. For example, the
495 cerebral cortex and thalamic areas have a higher mean **density** than the rest of the brain.

496 In a second stage, standardization is performed for $\mathbf{D}^{\text{gene normalized}}(T) = (d_{ij}^{\text{gene normalized}})$ in the
497 spatial domain, so that each gene in $\mathbf{T}$ has a distribution of gene expression densities with a
498 mean of 0 and a standard deviation of 1 over all sample positions

499 $\quad\quad d_{ij}^{\text{gene-space normalized}} = (d_{ij}^{\text{gene normalized}} - \mu_j) / \sigma_j \quad | \quad \forall\ d_{ij}^{\text{gene normalized}} \in \mathbf{D}^{\text{gene normalized}}(\mathbf{T}),$
500 $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad g_j \in \mathbf{T}$

501 where $\mu_j^{\text{gene normalized}} = \mu(d_{ik}^{\text{gene normalized}})_{k=1,\ldots,l})$ and $\sigma_j^{\text{gene normalized}} = \sigma(d_{ik}^{\text{gene normalized}})_{k=1,\ldots,l})$, d_{ik}
502 $\in \mathbf{D}(\mathbf{T}).$

503
504 <u>Effect calculation</u> is based on the trimmed mean of the gene-space normalized densitiy of all
505 genes in the function/trait set, that is called synergy $\mathbf{S} = (s(\mathbf{p}_i))_{i=1..n}$ where

506 $\quad\quad s(\mathbf{p}_i) = \mu_{\text{trimmed}}((d_{ik}^{\text{gene-space normalized}})_{k=1,\ldots,l}) \quad | \quad i=1,\ldots,n\ ,\ d_{ik} \in \mathbf{D}(\mathbf{T})$

507 With the synergy $\mathbf{S}$, several effects can be computed. Effects are divided into first order and
508 second order effects:

509   <u>First order effects</u> do not take the context of the network into account. The synergy $\mathbf{S}$ is a
510   first order effect itself, since $\mathbf{S}$ represents the gene function/trait association of every
511   sample point. Other first order effects would be the $\mu((d_{ik}^{\text{gene-space normalized}})_{k=1,\ldots,l})$ (which

15

512   is not robust to image artifacts like bubbles), or $\max((d_{ik}^{gene\text{-}space\ normalized})_{k=1,\ldots,l})$
513    $,((d_{ik}^{gene\text{-}space\ normalized})_{k=1,\ldots,l})$

514   Second order effects show the influence of the function/trait in the context of the
515   network. The function/trait-network association is expressed by weighting either
516   incoming or outgoing connections of every sample position by **S**, depending on the
517   scope of interest (afferent or efferent connections). Given a directed connectome as
518   connectivity matrix

519    $\mathbf{C} = (c_{vw})_{v,w = 1..n}$; $\mathbf{C} \in R^{n \times n}$

520   where the rows represent the source regions, the columns target regions, either an
521   incoming $\mathbf{C}^{\textbf{weighted in}}$ or outgoing $\mathbf{C}^{\textbf{weighted out}}$ weighted directed connectome is defined as

522    $\mathbf{C}^{\textbf{weighted out}} = (c_{vw}^{weighted\ out})_{w= 1,\ldots,n} = s(p_v) * (c_{vw})_{w = 1,\ldots,n} \quad | \quad \forall\ v = 1..n$

523    $\mathbf{C}^{\textbf{weighted in}} = (c_{vw}^{weighted\ in})_{v = 1,\ldots,n} = s(p_w) * (c_{vw})_{v = 1,\ldots,n} \quad | \quad \forall\ w = 1..n$

524   The second order effects on the network are computed by local network measures such as
525   incoming/outgoing node strength, hubs, authorities, closeness, betweenness and
526   eigencentrality on both incoming and outcoming weighted connectomes $\mathbf{C}^{weighted\ in}$ and
527    $\mathbf{C}^{weighted\ out}$. We showed in Fig. *2B*, that incoming/outgoing node strength performed best on
528   predicting our test data and is therefore stated exemplary. The incoming node strength (sum
529   of incoming connections for every node) of $\mathbf{C}^{weighted\ in}$ and $\mathbf{C}^{weighted\ out}$ is defined as

530    $\mathbf{IN}^{weighted\ out} = (in^{weighted\ out}(p_v))_{v=1..n}$ where
531    $in^{weighted\ out}(p_v) = \sum_{w=1}^{n} c_{vw}^{weighted\ out} \mid \forall\ v = 1..n$

532    $\mathbf{IN}^{weighted\ in} = (in^{weighted\ in}(p_v))_{v=1..n}$ where
533    $in^{weighted\ in}(p_v) = \sum_{w=1}^{n} c_{vw}^{weighted\ in} \mid \forall\ v = 1..n$

534   and the outgoing node strength (sum of outgoing connections for every node) as

535    $\mathbf{OUT}^{weighted\ out} = (out^{weighted\ out}(p_w))_{w=1..n}$ where
536    $out^{weighted\ out}(p_w) = \sum_{v=1}^{n} c_{vw}^{weighted\ out} \mid \forall\ w = 1..n$

537    $\mathbf{OUT}^{weighted\ in} = (out^{weighted\ in}(p_w))_{w=1..n}$ where
538    $out^{weighted\ in}(p_w) = \sum_{v=1}^{n} c_{vw}^{weighted\ in} \mid \forall\ w = 1..n$

539

540   Statistical evaluation of the computed effects (first and second order) are performed by
541   comparing them to the effects of random drawn gene sets (genome-wide randomized
542   function/trait-gene association) from the genome **G**.

543   1. Calculate the network effects for a function/trait **T**.
544   2. Draw >=1000 random set of genes from the genome **G** with equal size of **T**.
545   3. Calculate the first and second order effects for every random set.

Ganglberger et al.
Functional neuroanatomy from genes

546   4. P-values for the effects of **T** can be computed for every spatial sample position by
547   performing a t-test against the null-distribution represented by the >=1000
548   random effects since every spatial sample point is normally distributed in the gene
549   dimension (verified by KS tests).

550   The significance of $\mathbf{IN}^{\text{weighted out}}$ can be interpreted as nodes that are receiving from primary
551   expression sites (regions with high **S**), while $\mathbf{OUT}^{\text{weighted in}}$ shows regions projecting to
552   primary expression sites.   P-value calculations of $\mathbf{IN}^{\text{weighted in}}$ and $\mathbf{OUT}^{\text{weighted out}}$ are
553   numerically equal to the p-value calculation of **S** (for a node degree>0), since for those cases
554   the sum of incoming and outgoing connections are constant factors when compared to
555   random effects. We point this out to clarify the p-value calculation of $\mathbf{IN}^{\text{weighted in}}$ and
556   $\mathbf{OUT}^{\text{weighted out}}$ can be substituted by **S** for computational reasons.

557   $$\text{in}^{\text{weighted in}}(p_v) = \sum_{w=1}^{n} s(p_v) * c_{vw} = s(p_v) * \sum_{w=1}^{n} c_{vw} \quad | \quad \forall\, v = 1..n$$

558   $$\text{out}^{\text{weighted out}}(p_w) = \sum_{v=1}^{n} s(p_w) * c_{vw} = s(p_w) * \sum_{v=1}^{n} c_{vw} \quad | \quad \forall\, w = 1..n$$

559   Due to the multiple comparison problem, we adjust the FDR of the p-values of the effects by
560   the Benjamini-Hochberg (13) method. The results in this paper are all significant under a
561   FDR <5% (if not indicated otherwise).

562   Output is a p-value map (a p-value for every spatial sample point) for every effect. In this
563   paper, **S**, **IN, OUT** are used due to their fast computation, simplicity and biological
564   significance.

Code availability

566   The code for retrieving data (gene expression, mouse connectome) from the AMBA API
567   consists of a Matlab script whose single input parameter is a .csv with function/trait
568   information as a list of gene symbols and Entrez IDs. The main algorithm was implemented
569   as an R-script that uses the generated files (downloaded data from AMBA) of the Matlab
570   script to normalize, calculate and carry out a statistical evaluation to generate p-value maps
571   and structural network visualization for every testcase. The statistical evaluation, which was
572   randomized because of the extent of the computational task, is parallelized.

573   MATLAB- and R-codes will be made publically available under an open source license for
574   non-commercial use upon acceptance of the paper for publication.

Figure generation

576   Figures were generated with a R-script that will be provided on request. It uses the p-value
577   maps of the method to generate slice-views of different effects, heatmaps with statistical
578   measures of the effects and gene expression, clustered networks, csv-files with raw data and
579   precision-recall heatmaps (for data with ground truth).

580   Slice-views: Slice-views show 11 maximum intensity projections of 5 sagittal slices each
581   of a 132x80x114 voxel volume (which represents spatial sample positions) that shows
582   the left hemisphere of the mouse brain. Slice-views are used to visualize a log-scaled
583   mapping of first order p-values (of **S**), second order incoming node strength **IN** (regions
584   that are targets of first order regions) and second order **OUT** (regions projecting to first

17

585   order regions). At the bottom-right corner is a color-bar, indicating the minus $\log_{10}$-
586   scaled p-values, the threshold for false positive FDR (10% solid line, 5% dotted line).
587   Slice-views of all testcases can be found in Supplementary Data 3 Case 1-30*A, B, C*.

588   Heatmaps: Heatmaps in Supplementary Data 3 Case 1-30*D* and Supplementary Data 4
589   show the log-scaled p-values of first and second order effects as well as single gene
590   effects (gene expression density of a gene vs gene expression density of the genome) for
591   every significant region (a region that has at least one voxel with significant first or
592   second order effect). The regions are color-coded (on the left side) corresponding to the
593   AMBA, and given by their acronym on the right side. Similar information can be found
594   in the attached csv files (Supplementary Data 1) which contain the region-wise p-values
595   of first and second order effects.

596   Clustered network graphs: We clustered our test sets via hierarchical clustering with
597   Ward's Criterion (59) using the R function `hclust(*, "ward.D2")`. To ensure that
598   voxels with similar connections are within the same cluster, they are clustered by their
599   Pearson-correlation coefficient of their connectivity. To visualize the clusters, we plotted
600   a sagittally-projected heatmap of their combined p-value (minimum p-value of effects),
601   surrounded by labels. The connectivity between clusters is shown by the sum of
602   connectivity (normalized by injection volume) between the clustered regions given as
603   grey-scale. All graphs can be found in Fig. 2*A* and Supplementary Data 3 Case 1-30*E*.

604   F_1-score bar-chart: Based on available ground truth from the literature (Supplementary
605   Data 2), we calculated the F_1-score (45) based on the precision and recall for a binary
606   classification of ordered p-values. It doesn't take the true negative rate into account,
607   which is acceptable for the following reason: The literature-based ground truth is region
608   based. This means we can identify

609   ▪ true positives (a positive classified voxel within a region of the ground truth)
610   ▪ false positive (a positive classified voxel outside a region of the ground truth)

611   but not

612   ▪ true negative (a negative classified voxel outside a region of the ground truth),
613   since the total set of regions of the functional neuroanatomy are still unknown
614   ▪ false negatives (a negative classified voxel within the ground truth), since it is
615   possible that only a subset of the ground truth region is specific for functional
616   neuroanatomy.

617   For the calculation of the F_1-score, respectively precision and recall, the precision is
618   computed as the ratio of true positive voxels to the amount of positive voxels. For a
619   voxel-based recall, a false negative rate would be necessary, and so we used the region-
620   based recall, the ratio of positive classified regions to ground truth regions. We defined a
621   positive classfied region if at least 5% of the voxels of a region is positive (to account for
622   noise). P-value maps for the F_1-score bar chart were computed at 200 micron resolution
623   due to extensive computational network measures.

624

Technical resources

625

626 We used the Amazon elastic cloud computing service with an "r3.8xlarge" instance (32 cores,
627 244 GB RAM) (60). More than 100 GB RAM is recommended, 40 GB alone to hold the
628 connectivity matrix in the memory. Additional memory is needed for parallel processing
629 (approximately 3 GB per core). We tested the R-scripts with 30 cores. The computation uses
630 about 200 GB Ram and takes between 1 and 2 hours per testcase (depending on the amount
631 of genes in a set) to calculate the p-values for first and second order effects. The clustering
632 for the circle-graphs are also parallelized. Depending on the size of the significant areas,
633 clustering takes between 30 minutes to 3 hours.

General statistics

634

635 Unless indicated otherwise, data were tested for normality by Kolmogorov–Smirnov or
636 D'Agostino & Pearson tests at $\alpha < 0.05$ and analyzed non-parametrically if tests didn't pass.
637 Predicted functional neuroanatomy maps were compared to ground truth from fMRI using a
638 Spearman correlation of the $-\log_{10}$-scaled voxel-wise p-value of predicted nodes, set to
639 $p=10^{-3}$ for all $p<10^{-3}$, to BOLD heat responses of wt animals or differences in BOLD heat
640 responses in Cacna2d3 mutant vs. wt animals, respectively. To compensate for registration
641 errors between the AMBA reference space and fMRI data, these comparisons were
642 performed on volumes downsampled to 400 μm spatial resolution.

References

644 56. Hawrylycz MJ, et al. (2012) An anatomically comprehensive atlas of the adult human
645 brain transcriptome. *Nature* 489(7416):391–9.
646 57. Fonov V, et al. (2011) Unbiased average age-appropriate atlases for pediatric studies.
647 *Neuroimage* 54(1):313–327.
648 58. Glasser MF, et al. (2013) The minimal preprocessing pipelines for the Human
649 Connectome Project. *Neuroimage* 80:105–124.
650 59. Murtagh F, Legendre P (2011) Ward's Hierarchical Clustering Method: Clustering
651 Criterion and Agglomerative Algorithm. *arXiv Prepr arXiv11116285* (June):20.
652 60. Amazon EC2 instance types. Available at: https://aws.amazon.com/ec2/instance-types/
653 [Accessed December 6, 2015].
654

Supplementary Note 1

655

656 Investigating functional and structural brain network data and its analysis is an ongoing
657 challenge (61). Bullmore and Sporns (61) described the exploration of structural and
658 functional brain networks as a multi-stage approach, beginning with the separate creation of
659 structural and functional connectivity matrices based on anatomical parcellations. Network
660 measures, such as node degree, node strength, hubs, centrality, betweenness etc., indicate
661 network properties of interest when compared to equivalent measures of a population of
662 random networks (null-distribution). A local (region-wise) or global (Mantel-test) (62)
663 comparison reveals functional and structural correspondences of the networks.

664 The integration of genetic information facilitates insight into the influence on neuronal
665 activity and structural organization of the brain (1). French and Pavlidis (1) compared cortical
666 and subcortical regions of a rat connectome (63) and AMBA gene expression data (8) using
667 Spearman's rank correlation to show that brain regions with similar expression patterns have
668 more similar connectivity profiles. The similarities are close enough that a computational
669 model by Ji et al (64) could predict structural connectivity by gene expression profiles. 4048
670 genes with coronal spatial expression data were used as individual features in a sparse model
671 to obtain a predictive accuracy of 93% on anatomical parcellations. A follow up study proved
672 that this also works on mesoscale-resolution (voxels at 200 micron resolution) (65).

673 A combined approach of comparing structural connectivity, gene co-expression correlation
674 and functional networks was investigated by (3). Resting-state fMRI networks (default-mode,
675 salience, sensorimotor and visuospatial) were used as a starting point to identify functionally
676 related cortical regions in mice and humans. The strength fraction (scaled node strength of
677 gene co-expression networks) between those regions was significantly more similar than to
678 the remaining brain regions (tested by permutation tests). Genes that are related to the four
679 functional networks were identified by ranking them by their marginal influence on the
680 strength fraction. A gene co-expression matrix including only top-ranked genes was
681 compared to structural connectivity using the Mantel procedure (62) and were significant
682 compared to a sample of 10,000 random gene sets. (2) used Spearman's rank correlation
683 between node degree of structural connectivity and gene co-expression of gene sets related to
684 Gene Ontology groups (cellular composition and biological process) to assess how structural
685 connectivity is genetically driven. Connectivity related Gene Ontology groups were also used
686 by Fulcher and Fornito (66). They showed that the mean gene co-expression correlation of
687 groups related to biological processes are higher for connections involving structural "hubs"
688 (node degree over threshold) vs non-hubs indicates topological specializations of
689 interregional connections. Structural network hubs were also found to correspond to known
690 functional networks from the literature (4, 5). Compared to other studies (1–3, 66) which
691 used node strength or variations of it, Rubinov and Sporns (12) assessed other structural
692 network parameters, such as community structures, hierarchical modules, high-low cost sub-
693 networks etc.

694 An overview of related work and its modalities can be found in Supplementary Table 1.
695 Apart from Fakhry and Ji (65), who used high-resolution prediction, the studies cited were
696 computed on anatomically parcellated mouse brains (Richiardi and Altmann (3) also used
697 human data). Our approach was performed on 100-micron grid parcellation. In contrast to
698 Richiardi and Altmann (3), where functionally related gene sets were products of their
699 marginal influence on resting-state networks, we used functionally-linked gene sets as the

700   entry point of our method. Fulcher and Fornito, as well as French et al. (2, 66) showed the
701   influence of Gene Ontology groups of biological processes on structural networks, while our
702   approach utilized sets from gene association studies (database-mining, QTL analyses or
703   SNPs) and that can be directly linked to certain behavioral or mental features. Known
704   functional networks from the literature confirmed our results as well as the correlation with
705   resting state fMRI.

706   Comparing gene co-expression correlation to structural connectivity is a common approach
707   for assessing brain structures with genetic functionality (1–3, 64–66, 4, 5). The novelty in our
708   paradigm is weighting structural connectivity with functionally related, cumulative gene. It is
709   not only comparing networks, but it shows the direct effect of functionally related gene
710   expression on brain anatomy. Those effects were encountered by node strength, which we
711   proved to be a sufficient indicator, but also with various other network measures.

712

713   **Supplementary Note 2**

714   Pain sensation is biomedically one of the most important brain functions. While physiological
715   sensation is essential to protect the organism and to avoid harm, it is very often a result of
716   diseases or pathological/abnormal processes when the sensory information does not reflect
717   the factual danger from the environment. Pain gene sets from mice and human were taken
718   from literature and databases (Supplementary Data 2) (67, 68). pre-clustered or pre-assigned
719   to subcategories based on behavioral phenotype (nociception, analgesia, hypersensitivity) or
720   functional   annotations   (Gene   Ontology   (GO)),   calcium   signaling   =   calmodulin
721   binding+calcium ion transport associated genes related to pain processing). For the human
722   case we chose a metastudy combining SNPs associated with pain sensitivity or we extracted
723   subcategories (obtained using the DAVID platform based on functional annotation) from the
724   database for pain-related genes. We also used the Calcium signaling category as a set based
725   on evolutionary conserved pain genes. Importantly, the effector networks from most of these
726   gene sets could be linked to known pain-related areas in the brain (46, 48, 69, 70), but also
727   other regions such as piriform and entorhinal cortices, nucleus accumbens and VTA (Fig.
728   4*A*). Functional neuroanatomy maps from these gene sets, and the single gene Cacna2d3,
729   were also compared to fMRI pain responses of wt and mutant animals, respectively (50) (Fig.
730   3*A*). The maps derived from the gene sets (except nociception) were similar to the expected
731   pain network from the mouse fMRI (Fig. 3*A*). The Cacna2d3-dependent maps identified by
732   our method retraced Cacna2d3's functional genetic effects on pain processing in fMRI in
733   regions like striatum, olfactory areas, somatosensory cortex, hippocampus, hypothalamus,
734   paraventricular nucleus of thalamus (PVT) and basal ganglia. Similarly, for the human gene
735   sets (Fig. 4*B*), we obtained the brain regions known to be involved in pain processing,
736   including central grey, PVT, insular and somatosensory cortex, but also VTA – as in the
737   mouse case – or higher order associative cortices which are responsible for self-awareness
738   and conscious perception of pain.

739   Fear and anxiety-related genes were retrieved from JAX QTLs database (mouse) or from
740   literature (mouse and human) (71, 72), pre-assigned to behavioral phenotypes (startle
741   response, exploration, anxiety, depression and panic disorder). Again, the computed maps
742   (mouse and human) contained nodes with a fitting functional annotation, like fear-related
743   regions in the amygdalar complex, prefrontal cortex, thalamic or midbrain structures (73–78).
744   Moreover, the main nodes detected by our method are in line with their associated functional

21

745 subcategory, e.g. startle behavior was linked to insular cortex and PVT, while mental
746 disorders were linked to insular cortex, ACB and VTA (Fig. 4*A*). For the panic disorder
747 category, we can see differences in cortical regions identified for mouse and human. For
748 example, human data, unlike the mouse, lacks vmPFC, somatosensory or motor corices,
749 while we did not detect the visual and auditory cortex in the mouse brain (Fig. 4).

750 For autism-related genes, we retrieved 183 genes implicated in behavioral phenotypes in
751 mouse models of ASD and 739 autism-associated genes in humans from Autdb database
752 (79) and clustered the genes with DAVID (80), for further analysis, we chose functional
753 annotation categories that were the most relevant for ASD modeling: linked to behavior,
754 cognitive abilities, synaptic functions and cellular level processes. Similar to the other gene
755 sets, the computationally predicted maps contained nodes related to autistic brain function
756 (71, 81–88), except parietal cortex for the mouse brain (Fig. 3*A*) and, in the case of the
757 human brain the cerebellar areas were not identified.

758 To sum up, we were able to identify most of the known functionally involved brain regions
759 for all of the investigated categories based on mouse and human data. Additionally, for
760 different specific subcategories the method identified functionally relevant structures which
761 were found at the highest positions in rank-order lists. Taking together all the data, the
762 method can also be a useful tool for identifying novel functional targets, potentially involved
763 in traits linked to the genetic input. With this, we can bridge already known functional
764 systems using potential new -still unexplored - connections or even identify new functional
765 networks. For more detailed information please see Supplementary Data 1, 2, Fig. 3,
766 Supplementary Data 3 Case 11-30 (for mouse) and Supplementary Data 4 (for human).

	Focus	Spatial gene co-expression	Functional data	Fusing/comparing data with structural connectivity	Quantitative network analysis
Predicting functional neuroanatomical maps from fusing brain networks with genetic information	Mouse, whole brain, voxels at 100-micron resolution Human, whole brain, 415 regions	Cumulative effect (trimmed mean co-expression) of gene sets	Functional related gene sets from literature and gene association studies (QTL analyses) + Gene Ontology Known functional networks from the literature Resting-state fMRI (heat response)	Weighting structural/functional connectivity with functionally related, cumulative gene sets to show their effect on the network via network measures Comparing network nodes to known functional networks from the literature and fMRI data	Network measures: Node strength, Hub/Authority, Betweenness, Closeness, Eigencentrality Comparing network measures of functional weighted networks with empirical null distribution (random-functional weighted networks)
(Review) Complex brain networks: graph theoretical analysis of structural and functional systems (61)	various, but especially human	-	fMRI, electrophysiological techniques (EEG, MEG or MEA)	Network measures of interest and compare them to equivalent parameters of random networks	various (e.g. node degree, clustering coefficients, motifs, hubs, centrality, modularity)
Integrative analysis of the connectivity and gene expression atlases in the mouse brain (64)	Mouse, whole brain, 301 brain regions	4048 genes (with non-zero expression and available as coronal sections) Genes were used as individual features in the prediction model	-	Using gene co-expression data to predict discretized (binarized by threshold) structural connectivity	-
High-resolution prediction of mouse brain connectivity using gene expression patterns (65)	Mouse, whole brain, voxels at 200-micron resolution	4000 genes (with non-zero expression and available as coronal sections) Genes were used as individual features in the prediction model	-	Using gene co-expression data to predict discretized (binarized by threshold) structural connectivity	-
Relationships between gene expression and brain wiring in the adult rodent brain (1)	Rat connectome, 142 distinct regions of nearly half the brain volume Mouse, whole brain, 142 regions	Gene expression correlation of 17,530 genes (filtered by unexpressed genes)	-	Spearman's rank Correlation between gene expression and node degree of structural connectivity Mantel correlation of connectivity graph and gene co-expression networks Comparing connectivity between single genes (thresholding the structural connectivity with regions where single genes are expressing)	Node degree of structural connectivity

Correlated gene expression supports synchronous activity in brain networks (3)	Mouse, 1777 nodes in cortical regions Human, cortical regions	Pearson Correlation genome wide	4 resting-state fMRI networks (default-mode, salience sensorimotor and visuospatial)	Compare strength fraction (=network measure) of gene co-expression correlation within/outside functional networks Marginal influence of each gene on strength fraction Comparing Mantel correlation of connectivity graph and transcriptional similarity Comparing Mantel correlation of structural/functional connectivity	Comparing strength fraction (scaled node strength) of gene-co-expression networks inter vs intra functional networks
Wiring cost and topological participation of the mouse brain connectome (4)	Mouse, whole brain, 112 regions	Gene expression correlation of 3380 genes (with non-zero expression and available as coronal sections)	Known functional networks from the literature	Compared hubs to known functional networks from the literature Compared gene expression profiles (nodal participation) to network hubs	Network measures: Community structures, hierarchical modules, hubs, high-low cost subnetworks
Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectomes (5)	Human, 308 cortical regions	Gene expression correlation of 20,737 genes	Gene sets related to synaptic transmission, regulation of glutamatergic signaling and potassium ion channels	Compared gene expression profiles to network hubs, modular community structures and connection distance of structural covariance matrix by correlation	Network measures: Node degree and Closeness-Centrality
Large-Scale Analysis of Gene Expression and Connectivity in the Rodent Brain: Insights through Data Integration (2)	Rat connectome, 142 distinct regions of nearly half the brain volume Mouse, whole brain, 142 regions	Gene expression correlation of gene sets of cell types and biological process division (Gene Ontology) that are related to connectivity	Cell-type enriched gene sets, Gene sets of Gene ontology groups (limited to biological process division)	Spearman's rank Correlation between gene expression and node degree of structural connectivity of a (cell-type enriched or GO group) gene set compared to empirical-null distribution (resampled gene sets)	Node degree of structural connectivity
A transcriptional signature of hub connectivity in the mouse connectome (66)	Mouse, whole brain, 213 brain regions	Pearson Correlation genome wide Mean co-expression of functional groups of genes	31 distinct functional groups of genes from biological process division (Gene Ontology)	Comparing mean gene co-expression correlation of functional groups for structural connections involving hubs vs non-hubs	Defining structural "hubs" as nodes with a node degree > k

767

768   **Supplementary Table 2**: Outline of related work with focus on the quantitative analysis of networks that are either functional,

769   structural, derived from gene expression, or a combination thereof.

		MOUSE	HUMAN
Pain	Meta-study: Database	Functional classes according to the Pain database (67)	Common human genetic variants (89–91) + Functional analysis with DAVID (80)
	Meta-study: literature screen		SNPs (89)
	Experimental data (literature)	Neuronal-specific RNAi knock-down strategy in adult *Drosophila* (68) + Functional clustering	

Fear	Meta-study: Database	Functional classes from JAX database: QTLs (72)	
	Meta-study: literature screen	Multidisciplinary integration of human (71, 90) and mouse (71, 91) data	

Autism	Meta-study: Database	Collection of all genes connected to ASD in humans and relevant animal models (AutDB) (79) + Functional analysis (DAVID) (80)

770   **Supplementary Table 3.** Summary of mouse and human functional genetic data collection.

Supplementary References

773   61.   Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of
774   structural and functional systems. *Nat Rev Neurosci* 10(maRcH):186–198.
775   62.   Mantel N (1967) The detection of disease clustering and a generalized regression
776   approach. *Cancer Res* 27(2):209–220.
777   63.   Bota M, Dong H-W, Swanson LW (2005) Brain Architecture Management System.
778   *Neuroinformatics* 3(1):015–048.
779   64.   Ji S, Fakhry A, Deng H (2014) Integrative analysis of the connectivity and gene
780   expression atlases in the mouse brain. *Neuroimage* 84:245–253.
781   65.   Fakhry A, Ji S (2015) High-resolution prediction of mouse brain connectivity using gene
782   expression patterns. *Methods* 73(August):71–8.
783   66.   Fulcher BD, Fornito A (2016) A transcriptional signature of hub connectivity in the mouse
784   connectome. *Proc Natl Acad Sci U S A*:1513302113–.
785   67.   LaCroix-Fralish ML, Ledoux JB, Mogil JS (2007) The Pain Genes Database: An
786   interactive web browser of pain-related transgenic knockout studies. *Pain* 131(1-2).
787   68.   Neely GG, et al. (2012) Construction of a Global Pain Systems Network Highlights
788   Phospholipid Signaling as a Regulator of Heat Nociception. *PLoS Genet* 8(12).
789   69.   Tracey I (2008) Imaging pain. *Br J Anaesth* 101(1):32–39.
790   70.   Denis D, Marouf R, Rainville P, A B, Nguyen D (2015) Effects of insular stimulation on
791   thermal nociception. *Eur J Pain*. doi:10.1002/ejp.806.
792   71.   Santos M, D'Amico D, Dierssen M (2015) From neural to genetic substrates of panic

793   disorder: Insights from human and mouse studies. *Eur J Pharmacol* 759:127–141.

794   72.   Eppig JT, et al. (2015) The Mouse Genome Database (MGD): Facilitating mouse as a
795   model for human biology and disease. *Nucleic Acids Res* 43(D1):D726–D736.

796   73.   Katche C, Dorman G, Slipczuk L, Cammarota M, Medina JH (2013) Functional integrity
797   of the retrosplenial cortex is essential for rapid consolidation and recall of fear memory.
798   *Learn Mem* 20(4):170–3.

799   74.   Ferreira TL, Moreira KM, Ikeda DC, Bueno OFA, Oliveira MGM (2003) Effects of dorsal
800   striatum lesions in tone fear conditioning and contextual fear conditioning. *Brain Res*
801   987(1):17–24.

802   75.   Bradfield L a, McNally GP (2010) The role of nucleus accumbens shell in learning about
803   neutral versus excitatory stimuli during Pavlovian fear conditioning. *Learn Mem*
804   17(7):337–343.

805   76.   Garcia R, Vouimba RM, Baudry M, Thompson RF (1999) The amygdala modulates
806   prefrontal cortex activity relative to conditioned fear. *Nature* 402(6759):294–296.

807   77.   Schoenbaum G, Chiba a a, Gallagher M (1998) Orbitofrontal cortex and basolateral
808   amygdala encode expected outcomes during learning. *Nat Neurosci* 1(2):155–159.

809   78.   Morrison FG, Dias BG, Ressler KJ (2015) Extinction reverses olfactory fear-conditioned
810   increases in neuron number and glomerular size. *Proc Natl Acad Sci* 112(41):12846–
811   12851.

812   79.   Basu SN, Kollu R, Banerjee-Basu S (2009) AutDB: A gene reference resource for autism
813   research. *Nucleic Acids Res* 37(SUPPL. 1).

814   80.   Huang DW, Lempicki R a, Sherman BT (2009) Systematic and integrative analysis of
815   large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.

816   81.   Kennedy DP, Adolphs R (2012) The social brain in psychiatric and neurological disorders.
817   *Trends Cogn Sci* 16(11):559–572.

818   82.   Critchley HD, et al. (2000) The functional neuroanatomy of social behaviour: changes in
819   cerebral blood flow when people with autistic disorder process facial expressions. *Brain*
820   123 (Pt 1:2203–12.

821   83.   Sahin M, Sur M (2015) Genes, circuits, and precision therapies for autism and related
822   neurodevelopmental disorders. *Science (80-)* 350(6263):aab3897–.

823   84.   Bickart KC, et al. (2014) The amygdala as a hub in brain networks that support social life.
824   *Neuropsychologia* 63:235–248.

825   85.   McAlonan GM, et al. (2005) Mapping the brain in autism. A voxel-based MRI study of
826   volumetric differences and intercorrelations in autism. *Brain* 128(2):268–276.

827   86.   Zhan Y, et al. (2014) Deficient neuron-microglia signaling results in impaired functional
828   brain connectivity and social behavior. *Nat Neurosci* 17(3):400–406.

829   87.   Bourgeron T (2015) From the genetic architecture to synaptic plasticity in autism
830   spectrum disorder. *Nat Rev Neurosci* 16(9):551–563.

831   88.   Anomal R, et al. (2015) IMPAIRED PROCESSING IN THE PRIMARY AUDITORY
832   CORTEX OF AN ANIMAL MODEL OF AUTISM. *Front Syst Neurosci* 9(158).
833   doi:10.3389/fnsys.2015.00158.

834   89.   Foulkes T, Wood JN (2008) Pain genes. *PLoS Genet* 4(7).

835   90.   Hovatta I, Barlow C (2008) Molecular genetics of anxiety in mice and men. *Ann Med*
836   40(2):92–109.

837   91.   Barth A, et al. (2014) Analysis of quantitative trait loci in mice suggests a role of Enoph1
838   in stress reactivity. J Neurochem 128(6):807–817.