University of Basel

How complex analyses of large multidimensional datasets advance psychology – examples from large-scale studies on behavior, brain imaging, and genetics

A Cumulative Dissertation

Submitted to the Faculty of Psychology, University of Basel,

in partial fulfillment of the requirements for the degree of Doctor of philosophy

by

M.Sc. Tobias Egli

from Wildberg ZH, Switzerland

Basel, Switzerland

January 2018

First supervisor: Prof. Dr. med. Andreas Papassotiropoulos

Second supervisor: Prof. Dr. med. Dominique J.-F. de Quervain

Chairperson of the doctoral committee: Prof. Dr. Jens Gaab

Approved by the Faculty of Psychology

At the request of

Professor Dr. med. Andreas Papassotiropoulos

Professor Dr. med. Dominique J.-F. de Quervain

Basel, the _____

_____

Dean

Abstract

Psychology investigates the interplay of human mind, body, and its environment in health and disease. Fully understanding these complex interrelations requires comprehensive analyses across multiple modalities and multidimensional datasets. Large-scale analyses on complex datasets are the exception rather than the rule in current psychological research. At the same time, large and complex datasets are becoming increasingly available. This thesis points out benefits, challenges and adequate approaches for analyzing complex multidimensional datasets in psychology. We applied these approaches and analysis strategies in two studies. In the first publication, we reduced the dimensionality of brain activation during a working memory task based on data from a very large sample. We observed that a mainly parietally-centered brain network was associated with working memory performance and global measures of white matter integrity. In the second publication, we exhaustively assessed pairwise interaction effects of genetic markers onto epigenetic modifications of the genome. Such modifications are complex traits that can be influenced by the environment and in turn affect development and behavior. The lack of observed strong interaction effects in our study suggested that focusing on additive effects is a suitable approach for investigating the link between genetic markers and epigenetic modifications. Both studies demonstrate how psychological scientists can exploit large complex datasets by applying adequate research practices and methodologies. Further adopting these approaches will prepare psychological research for harnessing large and complex datasets, leading towards a better understanding of mental health and disease.

Table of contents

Figure index

Acknowledgments

I would like to thank my supervisors Professor Andreas Papassotiropoulos and Professor Dominique J.-F. de Quervain for giving me the opportunity to complete my PhD in their lab working on unique and exciting projects. The work that has led to this thesis has evolved me in many ways, both personally and professionally.

I express my particular gratitude towards my family, my friends, and colleagues for supporting me in many ways. I thank Dr. Annette Milnik for her profound support and patient assistance. I am grateful for learning a lot while working with her. I show my appreciation to Dr. David Coynel who was always ready for lending a helping hand. I thank MSc. Andreas Aeberhard, MSc. Eva Loos, and MSc. Bernhard Fehlmann for their good company. I thank Dr. Virginie Freytag, Dr. Christian Vogler, and Dr. Angela Heck for many fruitful discussions and helpful advices on many occasions.

Finally yet importantly, I want to thank my partner Sarah Wüst for supporting me and keeping me going on this journey.

Abbreviations

fMRI	Functional magnetic resonance imaging
DNA	Deoxyribonucleic acid
SNP	Single nucleotide polymorphism
HPC	High performance computing
FWER	Family-wise error rate
FDR	False discovery rate
LD	Linkage disequilibrium
GSEA	Gene-set enrichment analysis
DWI	Diffusion weighted imaging
DTI	Diffusion tensor imaging
GWAS	Genome-wide association analysis
PCA	Principal components analysis
EFA	Exploratory factor analysis
PC	Principal component
ICA	Independent component analysis
IC	Independent component

1. Introduction

During the last three decades, the development of new neuroimaging tools has greatly facilitated investigating neurobiological correlates of psychological processes (Poldrack & Farah, 2015). Accordingly, neuroscience constitutes an integral part of contemporary psychological research (Schwartz, Lilienfeld, Meca, & Sauvigné, 2016). Combining techniques from psychology and non-invasive neuroimaging with tools of molecular biology and genetics has yielded promising insights into the molecular underpinnings of human behavior, cognitive functioning, and psychiatric disorders (Freytag et al., 2017; Heck et al., 2014, 2017; Milnik et al., 2012; Papassotiropoulos et al., 2013; Vogler et al., 2014). The technologies and methods applied in these fields are currently advancing at a fast pace (Medland, Jahanshad, Neale, & Thompson, 2014; Poline, Breeze, & Frouin, 2015; van Horn & Toga, 2014) and yield ever-growing amounts of increasingly complex and voluminous data (Fan, Han, & Liu, 2014). Such large and complex datasets may result from methods including the comprehensive analysis of brain connectivity (Burns, Vogelstein, & Szalay, 2014; Van Essen et al., 2013), whole genome or whole exome sequencing (Gudbjartsson et al., 2015; Heck et al., 2017), or from sources such as electronic health records (Boland, Hripcsak, Shen, Chung, & Weng, 2017; Geraci et al., 2017), mobile devices (Schobel, Pryss, & Reichert, 2015; Torous, Kiang, Lorme, & Onnela, 2016), social media (Luhmann, 2017; Park et al., 2014), online games (McNab et al., 2015; Stafford & Dewar, 2014; Stafford & Haasnoot, 2017), web content mining (Landers, Brusso, Cavanaugh, & Collmus, 2016), or deep phenotyping (Loeffler et al., 2015). The amount of available large datasets is in addition increasing due to recent collaborative efforts for acquiring very large research samples

(Medland et al., 2014) and the growing volume of publicly available data (Ferguson, Nielson, Cragin, Bandrowski, & Martone, 2014).

Genetics and neuroscience are commonly considered data-intensive research fields (Lazar, 2016; van Horn & Toga, 2014). In contrast, other subfields of psychology have only recently started to conduct large-scale analyses (Harlow & Oswald, 2016). Most psychological scientists are therefore used to investigating rather small datasets (Chen & Wojcik, 2016; Cheung & Jak, 2016). Research in psychology should embrace the opportunities that arise from investigating large and complex datasets – as an essential complement to small-scale studies. This will require adopting research practices and methodologies that enable harnessing vast amounts of complex data (Cheung & Jak, 2016; Harlow & Oswald, 2016).

This doctoral thesis contributes to the research field of psychology, firstly by highlighting benefits of analyzing large multidimensional datasets, secondly by pointing out challenges that arise from investigating such data, and thirdly by presenting adequate approaches for facing these challenges. I describe these approaches from the perspective of the researcher in molecular psychology and neuroscience but they are similarly applicable to other subfields of psychology and other scientific fields. Two studies investigating brain activation networks and epistasis demonstrate how adequate informatics infrastructure, statistical methods, and data visualization have enabled analyzing large datasets and gaining knowledge from vast amounts of data:

- Egli, T., Coynel, D., Spalek, K., Fastenrath, M., Freytag, V., Heck, A., Loos, E., Auschra, B., Papassotiropoulos, A., de Quervain, D. J.-F. & Milnik, A. (2018).

Identification of two distinct working memory-related brain networks in healthy young adults. *eNeuro, in press*

- Egli, T., Vukojevic, V., Sengstag, T., Jacquot, M., Cabezón, R., Coynel, D., Freytag, V., Heck, A., Vogler, C., de Quervain, D. J.-F., Papassotiropoulos, A. & Milnik, A. (2017). Exhaustive search for epistatic effects on the human methylome. *Scientific Reports*, 7, 13669.

In the first publication "Identification of two distinct working memory-related brain networks in healthy young adults", we applied dimensionality reduction to brain activation measured from $N = 26'542$ voxels during a working memory task in a large sample of $N = 1'369$ subjects. We then associated the resulting brain activation networks with individual performances in the task. The analysis revealed that a parietally-centered network was robustly associated with working memory performance. I designed the experiment, conducted the analyses, and wrote the paper.

In the second publication "Exhaustive search for epistatic effects on the human methylome", we exhaustively assessed pairwise interaction effects of $N = 192'955$ genetic markers scattered across the whole genome onto $N = 395'431$ deoxyribonucleic acid (DNA) methylation sites across the whole methylome in $N = 533$ subjects. This analysis was computationally highly intensive and required rigorous methodological precautions to counteract spurious effects. We therefore conducted a full replication in an independent sample of $N = 319$ subjects. The exhaustive analysis showed that pairwise interactions of genetic markers robustly affected a very small number of DNA methylation sites. I acquired the data, conducted analyses relevant for interpreting the results, and wrote the paper.

2. Theoretical background

2.1 Large and complex datasets

The total amount of data generated per day is estimated at 2.5 quintillion (2.5×10^{18}) bytes (Monteith, Glenn, Geddes, & Bauer, 2015). Scientific data follows this general trend and is constantly growing in size and complexity (Ma & Zhu, 2013). In neuroimaging, the data volume has duplicated every 26 months since 1995 (van Horn & Toga, 2014). The brain activation data acquired for a single subject may typically consist of 50M data points (50'000 voxels × 1'000 time points) or more. Hence neuroimaging data is massive in volume and highly dimensional (Fan et al., 2014). Owing to the development of relatively inexpensive high-throughput measurements, the volume of data in human genetics is growing even more rapidly (Fan et al., 2014); it has doubled every six or seven months for several years now (Gelernter, 2015). The sequenced genome of a single individual comprises approximately three billion base pairs (Venter et al., 2001), a recent study has identified 20M single nucleotide polymorphisms (SNPs) in sequenced individuals (Gudbjartsson et al., 2015). The cheaper SNP arrays measure SNPs scattered across the genome in a lower resolution of approximately 1M SNPs per subject (Corvin, Craddock, & Sullivan, 2010). Accordingly, the datasets generated in human neuroimaging and genetics have both been termed big data (Landhuis, 2017) and the combined application of neuroimaging data plus genetic data has been referred to as "really big data" (van Horn & Toga, 2014, p. 325) or "big data squared" (Lazar, 2016, p. 61). The term big data is not unambiguously defined (Chen & Wojcik, 2016; Cheung & Jak, 2016) and the quantifiable amount of data that is referred to as "big" can differ by several orders of magnitude when compared between different fields, e.g. between psychology and tech industry (Yarkoni & Westfall, 2017).

Nonetheless, various descriptions agree that big data involves datasets that could not be handled within a tolerable amount of time using traditional hardware and software tools (Chen & Wojcik, 2016; Chen et al., 2013; Chen, Mao, Zhang, & Leung, 2014). Big data is furthermore inherently complex (Fan et al., 2014; Monteith et al., 2015) with regards to large numbers of observations *n* and/or variables *p*.

Investigating a large multidimensional dataset or even integrating multiple such datasets yields several potential benefits. Psychology investigates complex traits such as cognition, emotion, and psychiatric disorders (Gratten, Wray, Keller, & Visscher, 2014; Matheson, 2017; Papassotiropoulos & de Quervain, 2011, 2015; Vogler et al., 2014). These traits show neural substrates in distributed brain circuits (Eriksson, Vogel, Lansner, Bergström, & Nyberg, 2015; Geib, Stanley, Wing, Laurienti, & Cabeza, 2017; Goodkind et al., 2015; Minzenberg, Laird, Thelen, Carter, & Glahn, 2010; Pessoa, 2017) and have complex genetic backgrounds (Debette et al., 2015; Heck et al., 2014; Munafò & Flint, 2014; Papassotiropoulos & de Quervain, 2011; Sullivan & Posthuma, 2014; Vogler et al., 2014). Correspondingly, isolated analyses of a few variables will not suffice for understanding the function and dysfunction of a system as complex as the human mind and brain (Akil, Martone, & van Essen, 2011; Yarkoni, Poldrack, Van Essen, & Wager, 2010). Rather than measuring single data points, analyses of complex patterns are required, e.g. investigating a broad set of psychological measurements in place of single ones (Krapohl et al., 2016; Loeffler et al., 2015), unstructured rather than structured data (Bedi et al., 2015; Geraci et al., 2017), polygenic effects instead of single genetic markers (Sullivan & Posthuma, 2014), or brain networks rather than single voxels (Akil et al., 2011; Poldrack, 2012). The synthesis of information across many variables, dimensions (e.g. spatial and temporal), modalities (e.g. cognitive

measurements, functional/structural MRI, genetics, and epigenetics), paradigms, and psychological domains may lead towards a better understanding of the relationship between mind and brain (Akil et al., 2011; Bogdan et al., 2017; Lessov-Schlaggar, Rubin, & Schlaggar, 2016; Logothetis, 2008; Poldrack, 2012; Sejnowski, Churchland, & Movshon, 2014; Yarkoni et al., 2010). In addition to hypothesis testing, explorative analyses in complex datasets across large numbers of variables allow to identify unexpected patterns and to build new hypotheses (Chen & Wojcik, 2016; Holzinger, Dehmer, & Jurisica, 2014; Monteith et al., 2015; van Horn & Toga, 2014). Of note, analyzing datasets from large samples provides more precise answers (Spiegelhalter, 2014) that are more representative of the underlying population (Yarkoni & Westfall, 2017). Additionally, large sample sizes enable to identify and investigate exceptional cases from the sample that would be excluded as outliers in smaller samples (Monteith et al., 2015).

Analyzing datasets of large volume and high complexity involves significant challenges with regards to informatics infrastructure, statistical methodology, and interpretation of results (Fan et al., 2014). In the following sections, I will address prominent challenges from the perspective of psychological research and point out strategies for approaching them.

2.2 Adequate informatics infrastructure and data management

Scientific studies need to be conducted, analyzed, and reported as transparent, reproducible, and as little error-prone as possible (Munafò et al., 2017; National Academy of Sciences, 2009; Open Science Collaboration, 2015). Meeting these demands while analyzing large and complex datasets requires apt informatics

infrastructure (Poline et al., 2015), as well as an adequate data management plan (Chen & Wojcik, 2016; Goodman et al., 2014; van Horn & Toga, 2014). The following section introduces informatics infrastructure and data handling strategies that enable conducting large-scale analyses with methodological and statistical rigor. Using the term "analysis" in the following sections will include the overall process of acquiring data, processing data, and conducting statistical tests or estimations, which is also referred to as an analysis pipeline (Yarkoni & Westfall, 2017).

Using programming or scripting languages alleviates the analysis of large and complex datasets, firstly by automating manipulations and computations that are repeated many times, and secondly because it keeps the human input at a minimum, which is less error prone (if scripted correctly; Wilson et al., 2014). In the context of scientific analyses, it is generally advisable to write code in high-level languages (like e.g. R or Python) and only to use low-level languages (such as C or Fortran) if performance needs to be optimized (Wilson et al., 2014). The high-level languages R and Python are widely used in large-scale data analyses in many research fields (Chen & Wojcik, 2016; Chen et al., 2014; Cheung & Jak, 2016). R and Python are open source languages with large communities of users and developers who contribute to an abundance of packages and libraries in many areas of application (Godsey, 2017; R Core Team, 2013; van Rossum, 1995).

Scripting complex analyses is likely to involve complicated codes that are difficult to oversee – possibly distributed across numerous sub-scripts. Version control tools like Git (https://git-scm.com) register changes in scripts and archive the different versions. This allows reverting all scripts to earlier versions if needed (Blischak, Davenport, & Wilson, 2016; Ram, 2013). Using Git also facilitates collaborative work on

complex tasks and analyses; on the one hand because it enables effortless sharing of scripts with collaborators or between different computers, on the other hand because it allows merging simultaneous changes by several individuals in the same script (Blischak et al., 2016; Ram, 2013). If a single script (that executes other subscripts) consolidates all processing steps and calculations applied in an analysis, it seamlessly documents the analysis. Importantly, this maximizes the transparency of the analysis, as it allows publishing the workflow alongside the manuscript of a paper and makes the entire analysis pipeline fully reproducible (Goodman et al., 2014; Nosek, Spies, & Motyl, 2012; Poldrack et al., 2017; Wilson et al., 2017). On Unix-like computer systems, scripts written in a Unix shell – e.g. the Bourne-again shell 'bash' – are beneficial for consolidating an analysis pipeline. This is especially helpful if the analysis comprises scripts written in several languages and/or tools executed from the Unix shell (Wilson et al., 2017). Because large-scale data processing and analyses require considerable amounts of memory and computational power (Fan et al., 2014; Medland et al., 2014; Poline et al., 2015), they can easily overburden individual desktop computers. In such cases, high performance computing (HPC) systems (computational clusters or supercomputers) may provide the required memory resources and processing power (Bouchard et al., 2016; van Horn & Toga, 2014). Even if a desktop computer could handle the memory and provide the computational power required for an analysis, using an HPC system may speed up calculations considerably (Godsey, 2017). In order to use such systems efficiently, researchers profit greatly from collaborating with facilities dedicated to HPC and data storage or at least from interacting closely with computer scientists and informaticians (Bouchard et al., 2016; Cheung & Jak, 2016; Poldrack, 2012; van Horn & Toga, 2014).

The value of a scientific analysis depends on its reproducibility (Holzinger et al., 2014; Kleppner & Sharp, 2009). Reproducible analyses require the ability to trace back all the data investigated and each processing step applied during an analysis (Goodman et al., 2014; Wilson et al., 2014). Therefore, the data that serves as the starting point of an analysis should be stored as "pure" and unprocessed as possible (Hart et al., 2016). Storing this "raw data" with read-only access permission prevents unwanted manipulations of the data (Wilson et al., 2017). Furthermore, storing the raw data redundantly in several locations and using various storage systems prevents data loss (Berman, 2008). The integrity of stored data can be monitored by saving the cryptographic hash (e.g. SHA or MD5) of each dataset as metadata (Hart et al., 2016); any silent corruption and/or manipulation of a dataset will change the associated cryptographic hash. In order to keep track of the data's location and state, the cryptographic hash, the path to the data in the storage system, and other metadata should be systematically saved for each dataset (Berman, 2008). Accessing data contents for further processing or for conducting an analysis is then possible via the path to the data that is stored in the metadata. Figure 1 illustrates a schematic example of a scientific data management plan that reassures traceability of the data as well as the applied processing steps and analyses.

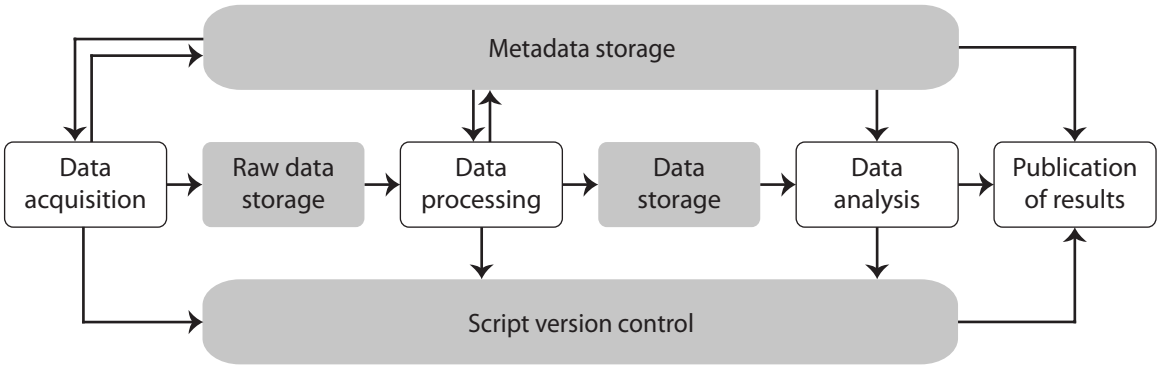

Figure 1. Schematic representation of an exemplified data management plan. Grey boxes represent storages of data or metadata; white boxes depict operations on the data. An arrow towards a storage location represents storing new data; an arrow from a storage location represents accessing stored data.

During my PhD studies, I have participated in planning, designing, and developing various automated data analysis pipelines, or components of such pipelines, respectively. I was particularly involved in the validation of raw data, storing the raw data in a secure file system, and storing the associated metadata and/or data contents in a scientific data warehouse based on LabKey (Nelson et al., 2011) and HDF5 (http://www.hdfgroup.org/HDF5). I also prepared the raw data of a behavioral n-back task for statistical analyses (which was used in Egli et al., 2018), including outlier detection and data aggregation. I furthermore developed an analysis pipeline for extensive dimensionality reduction of functional brain imaging data in Egli et al. (2018). In Egli et al. (2017), I participated in an analysis that used graphics processing units of an HPC environment for efficiently parallelizing quadrillions of computations using the software EpiGPU (Hemani, Theocharidis, Wei, & Haley, 2011).

2.3 Statistical challenges

Datasets investigated in molecular psychology and in neuroscience often comprise large numbers of variables *p* and fewer observations *n* (Lazar, 2016). This "small n large p problem" (Spiegelhalter, 2014, p. 264) is typically encountered in functional magnetic resonance imaging (fMRI) studies with samples of a few hundred or thousand individuals, which measure brain scans across millions of voxels (in our brain imaging study ~50'000 × 1'000 voxels in 1'400 subjects). Similarly, the problem also occurs in genetic and epigenetic studies, which measure hundreds and thousands of genetic and epigenetic markers (in our epistasis study ~190'000 SNPs × 400'000 CpGs in 500 subjects). In the following sections, I outline various statistical challenges that arise from such data characteristics and that we encountered in our studies. I also point out how we approached these challenges.

2.3.1 Multiple comparisons

In Egli et al. (2018), we conducted hypothesis tests across large numbers of voxels, and in Egli et al. (2017) we computed enormous numbers of interaction analyses. Conducting large numbers of statistical hypothesis tests extensively accumulates false-positive results (Poldrack et al., 2017). The probability of making any false-positive inference by a group or family of tests is termed family-wise error rate (FWER); the proportion of false-positive results that is expected among all tests is referred to as false discovery rate (FDR; Cao & Zhang, 2014). FWER corrections, e.g. Bonferroni adjustment or Westfall-Young permutation (Westfall & Young, 1993), are rather stringent and potentially lead to a lower detection rate for true effects (Cao & Zhang, 2014). In contrast, FDR corrections like the Benjamini-Hochberg method tolerate a

minor amount of false-positive results in order to improve the chance of detecting true effects (Benjamini & Hochberg, 1995). Therefore, the decision whether to correct for FWER or for FDR should trade off the benefits and drawbacks of false-positive and false-negative observations. In exploratory analyses that involve large numbers of tests and expect many true negative results, FDR correction is more suitable (Glickman, Rao, & Schultz, 2014). In contrast, FWER correction is more appropriate for confirmatory analyses (Frane, 2016). We accordingly corrected for FDR when associating working memory brain activation (across all voxels as well as across estimated brain networks) with task performance measures in Egli et al. (2018). Due to its more complex algorithm, FDR correction has the disadvantage of increased computational demands, when compared to FWER correction. In light of the enormous number of computations, we therefore applied FWER corrections instead of the statistically more suitable FDR corrections to account for 7.36×10^{15} epistasis tests in Egli et al. (2017) to circumvent unnecessary computational burden and complexity.

As an alternative to correcting for multiple comparisons, reducing the dimensionality of the investigated dataset can increase the sensitivity and the efficiency of analyses on complex datasets (Medland et al., 2014). In Egli et al. (2018), we accordingly applied dimensionality reduction to brain activation (in addition to analyses across all variables). This reduced the dataset from 26'542 voxels to six brain activation networks and facilitated detecting associations of brain activation with other measurements.

2.3.2 Correlated variables

Correcting for the total number of comparisons may be too restrictive if the tested variables are highly correlated (Poline et al., 2015). This is generally the case when analyzing genetic markers in linkage disequilibrium (LD; Wray, 2005), or fMRI signals in neighboring voxels (Medland et al., 2014). In Egli et al. (2017) we only included uncorrelated genetic markers in the analysis. By contrast, the dimensionality reduction applied in Egli et al. (2018) yielded a low number of statistically independent and uncorrelated features of brain activation. Both approaches are suitable for circumventing issues related to highly correlated variables.

2.3.3 Spurious associations and replications

In both studies, we conducted exploratory hypothesis tests across large numbers of variables. Exploratory tests in complex datasets are prone to spurious results (Button et al., 2013; Ioannidis, 2005; Szucs & Ioannidis, 2017), and therefore require measures for counteracting false-positive findings. The best method for validating promising findings is replication in independent samples (Bogdan et al., 2017; Medland et al., 2014; Nosek et al., 2012; Yarkoni et al., 2010), especially in exploratory analyses (Poldrack et al., 2017). In Egli et al. (2017), we fully replicated the findings of our epistasis analyses in an independent sample. If a replication based on independent data is not feasible, other appropriate validation methods include within-sample cross-validation, meta-analytical approaches, evaluation of convergence across methods, or conceptual replication (Bogdan et al., 2017; Nosek et al., 2012; Yarkoni & Westfall, 2017). Because we lacked a replication sample with brain imaging measurements during working memory performance, we compared our findings to meta-analytic

results, and applied within-sample cross-validation as well as resampling for validating our working memory brain activation networks in Egli et al. (2018).

2.3.4 Computational challenges

Elaborate statistical methods, such as the dimensionality reduction in Egli et al. (2018), or the exhaustive search for epistatic effects in Egli et al. (2017), are computationally very expensive when applied to large numbers of variables and observations. Sequentially aggregating the raw data to summary statistics on different levels of the data is one option for efficiently solving this problem. In the field of fMRI, this procedure is referred to as level-wise analysis (Holmes & Friston, 1997). In Egli et al. (2018), we calculated for each subject separately the summary statistics that describe working memory-related brain activation in each individual voxel (first-level statistics). We then applied dimensionality reduction to these values on the group level (i.e. across all subjects; second-level statistics). Such computationally efficient split-apply-combine approaches allow parallel computing of the apply-step (Cheung & Jak, 2016; Kane, Emerson, & Weston, 2013). This approach is especially suitable when analyzing hierarchically structured datasets like repeated measurements, where data points within subjects represent the level 1 units and the individuals represent the level 2 units (Goldstein, 2011). In our epistasis analysis in Egli et al. (2017), we followed a different approach and gained computational efficiency by applying a simplified and computationally less demanding analysis strategy (as suggested in Wei, Hemani, & Haley, 2014). We used EpiGPU (Hemani et al., 2011) for computing the exhaustive $N = 7.36 \times 10^{15}$ calculations across $N = 1.85 \times 10^{10}$ pairs of SNPs and $N = 395'431$ CpG sites. EpiGPU is computationally very efficient, but merely approximates a true interaction

test. For all SNP-SNP-CpG combinations that were indicative of an interaction effect in this screening step, we then calculated statistically more appropriate linear regressions for confirming the interaction results. The screening based on a simplified analysis strategy effectively reduced this computationally costly analysis step to $N = 9.54 \times 10^9$ calculations, which is merely 0.00013% of the original number of computations.

2.4 Interpreting results

The wealth of information generated from analyzing complex datasets can be difficult to absorb, understand, and interpret (Sejnowski et al., 2014). Methods that can alleviate these issues include data visualization, annotation of data with additional information, or third level statistics that combine the outcome of multiple statistical analyses. Applying these approaches can help to gain further insights from the derived results.

2.4.1 Data visualization

Comprehending higher dimensional datasets tends to overburden human perception. In such cases visualizations may help to map data into lower dimensional space (Holzinger et al., 2014). Modern data illustrations are not merely interchangeable with statistical tables but provide additional qualities like integrating multidimensional data from different sources (Tufte, 2001). Accordingly, data visualization can give a sense of relations in data that were not intelligible in any other way (Fox & Hendler, 2011). Visualizations are therefore critical for understanding complex data. However, designing the appropriate visualization for a given dataset is not an easily performed method, but should rather be regarded a form of art and expert storytelling (Fox &

Hendler, 2011; Murray, 2013). Edward R. Tufte, a pioneer in the field of data visualization (Unwin, 2008), has described graphical excellence as a matter of both statistics and design that is almost always multivariate, illustrates data as comprehensive as possible, and uses only as much elements in the illustration as necessary (Tufte, 2001). The concept of visual data fusion is helpful for illustrating complex data as it integrates data from different modalities into a single visualization based on a common frame of reference (Kehrer & Hauser, 2013). The common reference allows comparative visualization displays, which depict differences and similarities in the data by juxtaposition, overlaying, or plotting of computed relationships (Kehrer & Hauser, 2013).

Various software applications provide excellent default displays of commonly used illustration types (Deepayan, 2008; Wickham, 2009). Other tools produce illustrations that are specific for particular research fields, for instance the Python library 'PySurfer' (https://pysurfer.github.io) or the standalone program 'MRIcroGL' (http://www.mccauslandcenter.sc.edu/mricrogl) for brain imaging. Such tools provide graphics that are standard in the field and are easily understood by readers familiar with them (Unwin, 2008). I used MRIcroGL for visualizing functional brain networks as three-dimensional renderings in a semi-transparent brain in Egli et al. (2018), see Figure 2a. These visualizations allowed perceiving entire brain networks based on a low number of images. The classically used "brain slices" can yield good visualizations of individual regions of interest, but cannot easily convey more complex patterns of multiple regions that are distributed across the brain, as is illustrated in Figure 2b.

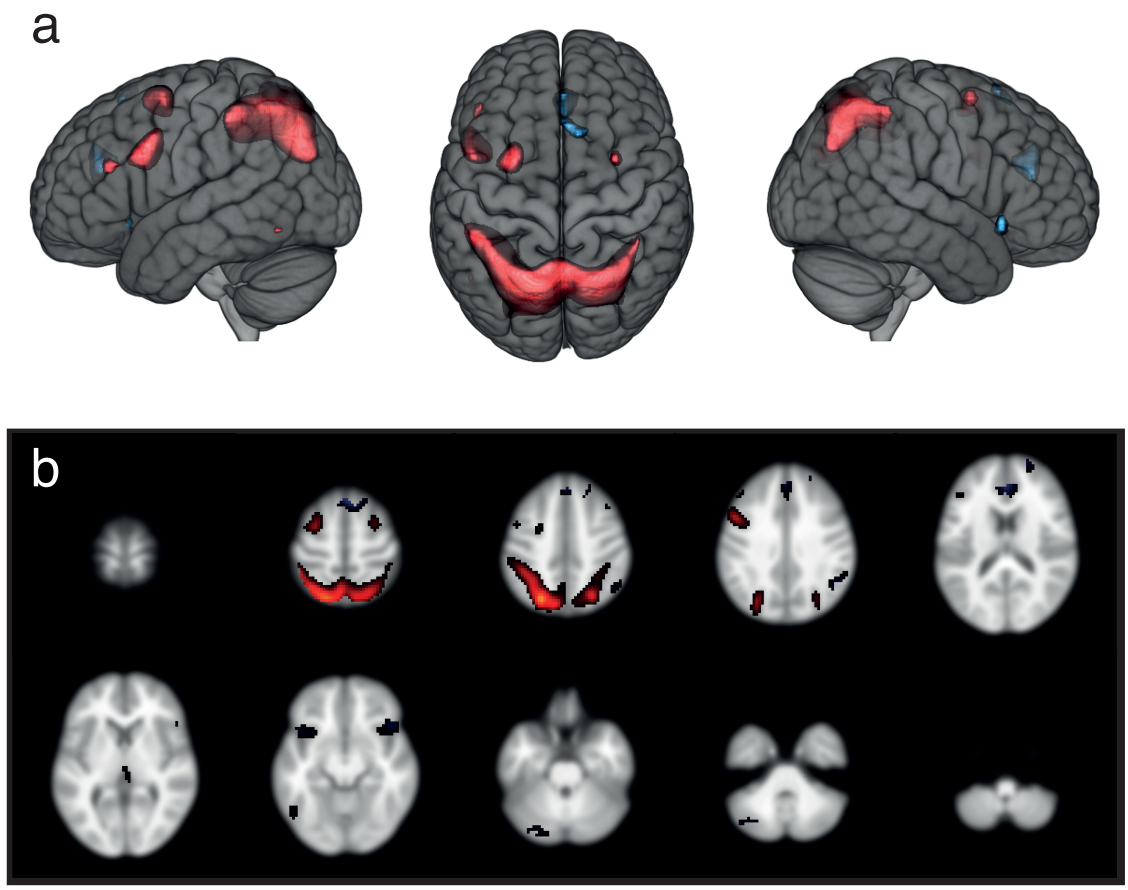

Figure 2. Different illustrations of a distributed brain network. (a) Three-dimensional renderings of the brain network in a semi-transparent brain (left lateral, superior, and right lateral views), created using MRIcroGL. (b) The same brain network represented in ten horizontal slices of the brain (from top to bottom), created using the R-package 'grid'.

In Egli et al. (2017), we visualized the local functional backgrounds of genomic regions associated with our results. These illustrations used visual data fusion as well as comparative visualization displays for integrating the genetic and epigenetic markers with additional information, either derived from the investigated data or retrieved from the UCSC genome browser (Tyner et al., 2017), see Figure 3.

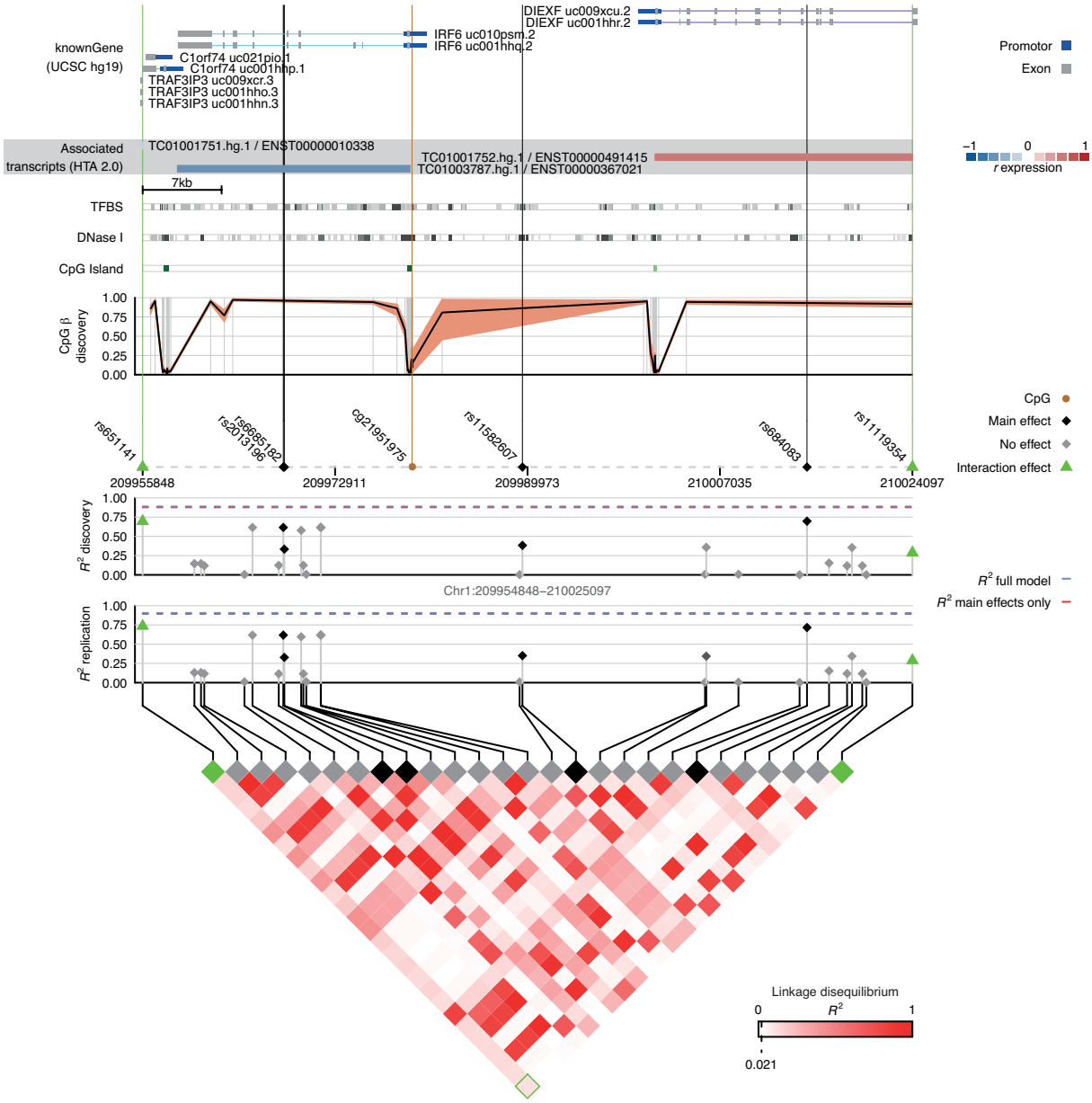

Figure 3. Example of a multivariate figure created using the R-package 'grid' showing an interaction effect and several main effects of SNPs onto DNA methylation in one CpG-site. Visual data fusion (different horizontal panels with concordant x-axes) integrates the data from different sources (external data retrieved from the UCSC genome browser; Tyner et al., 2017). Comparative visualization (vertical lines) allows comparing information between the different sources. From Egli, et al. (2017), supplementary, licensed under Creative Commons CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Because no software applications were available for creating such visualizations per default, I developed a tool for creating the illustrations using the R-package 'grid' (Murell, 2006). The package 'grid' does not contain high-level functions for producing complete illustrations but provides low-level graphics functions. Low-level functions give the user extensive control over all aspects of the illustration, but also require more expertise in coding as compared to high-level functions (Unwin, 2008).

2.4.2 Integrating results with additional information

Combining new results with findings from past studies further improves the interpretability of new findings (Yarkoni et al., 2010). The comparison with former studies also allows assessing the plausibility of new results (Woo, Chang, Lindquist, & Wager, 2017). Recently introduced resources provide large collections of results from past studies in neuroimaging (NeuroSynth; Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011) or in genetics (NHGRI-EBI GWAS Catalog; Macarthur et al., 2017) that can be used for this purpose. In Egli et al. (2018), we identified a working memory-related brain activation network. By using the meta-analytic results from 11'406 fMRI studies in NeuroSynth, we showed that the spatial characteristics of this network had also been observed across a large number of other brain imaging studies. In addition to results from former studies, information from expert-curated databases can provide annotations for contextualizing new results. Corresponding databases of genetic information include dbSNP (Kitts, Phan, Ward, & Holmes, 2014), UCSC Genome Browser (Tyner et al., 2017), Gene Ontology (The Gene Ontology Consortium, 2013), Reactome (Haw, Hermjakob, D'Eustachio, & Stein, 2011), or the Kyoto Encyclopedia of

Genes and Genomes (Kanehisa et al., 2014). In Egli et al. (2017), we used information from the UCSC Genome Browser for visually annotating our results (see Figure 3).

Besides data visualization and descriptive annotations, statistical methods can add informative value to new findings from complex analyses that are difficult to oversee (Pers, 2016). For instance gene-set enrichment analysis (GSEA) adds biological context to findings from genetic analyses (Mooney & Wilmot, 2015). GSEA either tests whether a group of genes or genetic markers (e.g. with an association in a genome-wide association study above a certain threshold) significantly overlaps with a predefined set of genes (Mooney & Wilmot, 2015; Pers, 2016), or whether the genes in a gene-set are jointly associated with a given trait (Wang, Li, & Hakonarson, 2010). In Egli et al. (2017), we used GSEA for assessing functional commonalities of our main results, the epigenetic modifications affected by epistasis. The epigenetic markers overlapped with gene-sets implicated in HPV infection as well as cancer. While this finding was not essentially associated with our research question, it added some plausibility to our findings, since genome-wide epistasis analyses had also shown small numbers of epistatic effects on cancer risk (Shen, Li, Song, Chen, & Shi, 2017).

3. Methods

3.1 Neuroimaging

Over the last two decades, magnetic resonance imaging (MRI) has evolved into one of the most applied non-invasive methods in neuroscience (Fan et al., 2014; Poldrack & Farah, 2015). MRI infers three-dimensional measurements of brain structures and brain activation from the spin of nuclei (commonly hydrogen atoms; Logothetis, 2008). The most frequent MRI techniques include structural MRI, functional MRI (fMRI), and

diffusion weighted imaging (DWI) or diffusion tensor imaging (DTI), respectively. Structural MRI makes use of different signal properties in distinct tissue types for segmenting the brain into cortical and subcortical structures, white brain matter, or cerebrospinal fluid (Desikan et al., 2006). Instead of structural variation, fMRI infers changes in neuronal activity from variations in the oxygenation of hemoglobin. The deoxygenation observed in a brain area is interpreted as higher consumption of oxygen in that region, which is in turn assumed as a proxy for higher brain activation (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001). DWI estimates the diffusion of molecules (mainly water) in tissues (Jones, Knösche, & Turner, 2013). DTI, a subtype of DWI, allows measuring the diffusion in neuronal tracts, it is therefore extensively used for characterizing white matter tracts (Beaulieu, 2002). In Egli et al. (2018), we used fMRI for estimating functional brain networks and associated them with individual task performances as well as with white matter properties measured using DTI.

3.2 Genetic and epigenetic analyses

Genetic variation results from differences in the sequence of nucleic acids in the DNA. Loci in the genome with differing single nucleic acid pairs (alleles) that are common in the population (e.g. present in at least 1%) are termed single nucleotide polymorphisms (SNPs; Poline et al., 2015). For regulating the transcription of genes, transcription factors physically bind to the DNA. Chemical modifications on the DNA or its surrounding regions can therefore impact the transcription factors' ability to access the DNA (Zhang & Meaney, 2010). The molecular processes that impact gene transcription without altering the sequence of nucleotides are termed epigenetic

events or marks (Bird, 2007). DNA methylation is a classical epigenetic alteration that (in mammals) adds a methyl group onto CpG dinucleotides in the DNA (Li, 2002). DNA methylation can silence gene transcription by preventing transcription factors from binding to the DNA (Bird, 2002). Alternatively, it can indirectly increase the transcription by silencing genes that involve acetylation of histone proteins. The DNA sequence is spooled around the histone proteins. Reduced acetylation of histones results in less dense packing of the DNA and consequently making the DNA more accessible for transcription (Klose & Bird, 2006). Evidence suggests that environmental events impact epigenetic marks in early life, which in turn influences neural development and ultimately brain function as well as behavior (Weaver et al., 2004; Zhang & Meaney, 2010).

Microarrays allow measuring genetic or epigenetic markers scattered across the genome with a relatively low resolution of ~1M SNPs per subject (Corvin et al., 2010) or ~450'000 CpG-sites per subject, respectively (Bibikova et al., 2011). Genome-wide association studies (GWAS) associate each individual SNP with a given trait (e.g. using chi-squared test, linear regression, or logistic regression; Corvin et al., 2010). When used for investigating complex traits, GWAS typically yield risk variants with small effect sizes (Gelernter, 2015; Papassotiropoulos & de Quervain, 2015; Poldrack et al., 2017). Correspondingly, the proportion of variation in complex and polygenic traits that is explained by additive effects of all significantly associated SNPs is usually low (it typically sums up to less than 10%; Visscher, Brown, McCarthy, & Yang, 2012). It is therefore speculated that some of the remaining variation (the "hidden variance") could be explained by investigating non-additive effects, for instance in epistasis analyses that investigate interaction effects between SNPs (Wei et al., 2014). In Egli et

al. (2017), we assessed pairwise SNP-SNP interaction effects onto the methylation of CpG sites across the genome and the methylome, both measured from microarrays. For genetic or epigenetic analyses that yield large amounts of results, it may be beneficial to assess common biological and/or functional implications of the results.

3.3 Dimensionality reduction

Dimensionality reduction techniques aim at reducing the dimensionality of the data while retaining as much of the relevant information as possible (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Kehrer & Hauser, 2013). Such techniques commonly either select a subset of features in the data (without transforming the data), or construct new features from the data, for instance describing linear combinations of the variables in the data (Ma & Zhu, 2013; Mladenić, 2006). Classical psychological research frequently transforms data to some low dimensional representation using principal component analysis (PCA) or exploratory factor analysis (EFA; Fabrigar et al., 1999). PCA estimates new variables that describe the main sources of variance in a dataset (Jolliffe, 2002). It successively estimates principal components (PCs) that account for as much variance in the data as possible, are orthogonal to the preceding component, and are uncorrelated. Accordingly, the first PC explains the most variance; the second PC explains the most of the remaining variance, etc. Other than PCA, EFA aims at revealing a predefined number of latent variables that underlie the covariation of the observed variables (Fabrigar et al., 1999). Studies in the fields of neuroimaging, molecular biology, and genetics increasingly apply independent component analysis (ICA) or penalized regression for dimensionality reduction (Kong, Vanderburg, Gunshin, Rogers, & Huang, 2008; Medland et al., 2014). While PCA and EFA rely on

Gaussian signals and the covariance of variables, ICA takes into account the total dependence structure of all variables and uses higher-order statistics to find a linear representation of non-Gaussian data (Hyvärinen & Oja, 2000; Kong et al., 2008). Like the more classical methods of EFA and PCA, ICA results in a linear decomposition of the data. It is additionally able to separate sources that are mixed in the observed data, a task where the classical methods frequently fail (Hyvärinen, 2013). The separation of mixed signals by ICA is illustrated and compared to PCA in Figure 4.

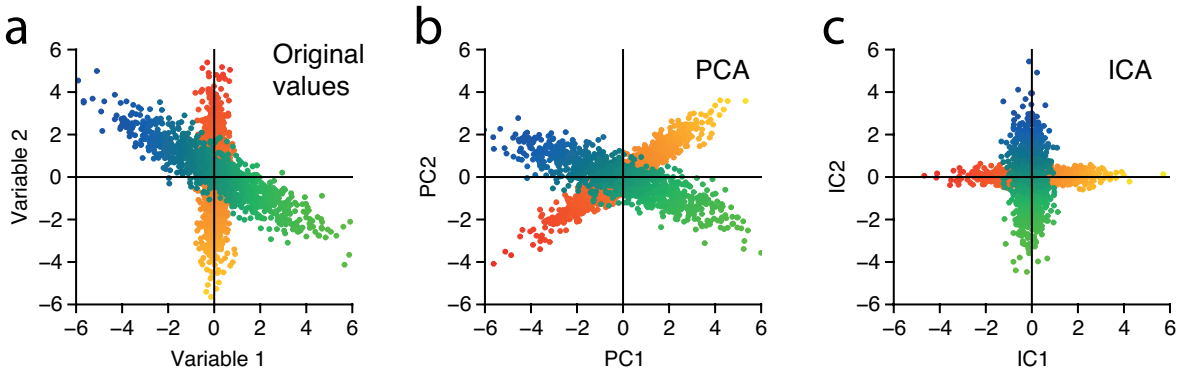

Figure 4. The decompositions of two mixed signals using PCA and ICA. (a) A mixture of two signals (blue-green and red-yellow) measured by two variables. (b) The PCs from a PCA explain the main sources of variance but are not able to disentangle the mixed signals. (c) Conversely, the ICA yields statistically independent estimates of the two signals.

In contrast to the results of PCA and EFA, the independent components (ICs) estimated by ICA are not only uncorrelated, but also as statistically independent as possible. Figure 5 depicts examples of variables that are uncorrelated and either statistically independent or dependent from each other.

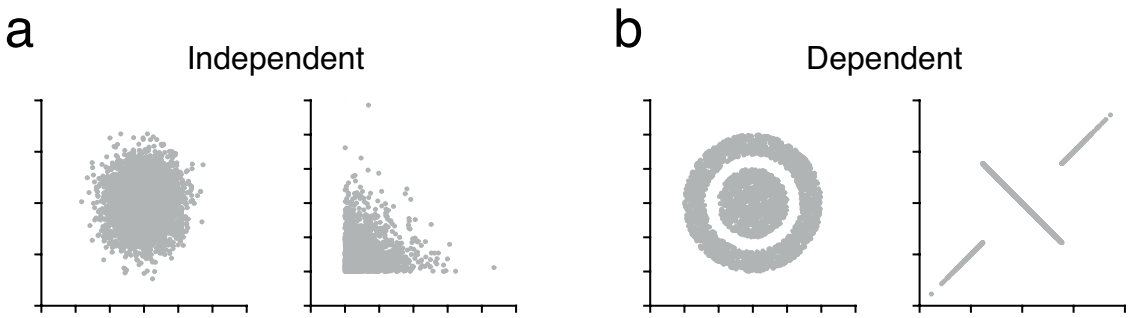

Figure 5. (a) Two examples of uncorrelated and statistically independent variables. (b) Two examples of uncorrelated and statistically dependent variables.

In Egli et al. (2018), we used ICA to reduce the dimensionality of brain activation during a working memory task and retrieved networks of statistically independent brain activation. We applied the ICA decomposition using a very efficient implementation of ICA in the R-package 'fastICA' (Hyvärinen & Oja, 2000). Applied to the matrix X of m observations (1'369 subjects) across n variables (working memory brain activation in 26'542 voxels), the ICA estimated a matrix S of k × n latent sources (6 voxel-wise loadings) that underlay the variables while holding the voxel loadings as statistically independent from each other as possible (Engreitz, Daigle Jr., Marshall, & Altman, 2010). In addition to the voxel loadings, ICA also yielded a matrix A of m × k mixing coefficients (subject-wise scores) for each IC. The mixing coefficients of a particular component depicted the projection of the original brain activation data onto this component's estimated voxel loadings, such that X = AS (Hyvärinen, 2013; Hyvärinen & Oja, 2000). Figure 6 summarizes the estimates involved in ICA; Figure 7 visualizes the voxel loadings that resulted from the ICA on working memory brain activation.

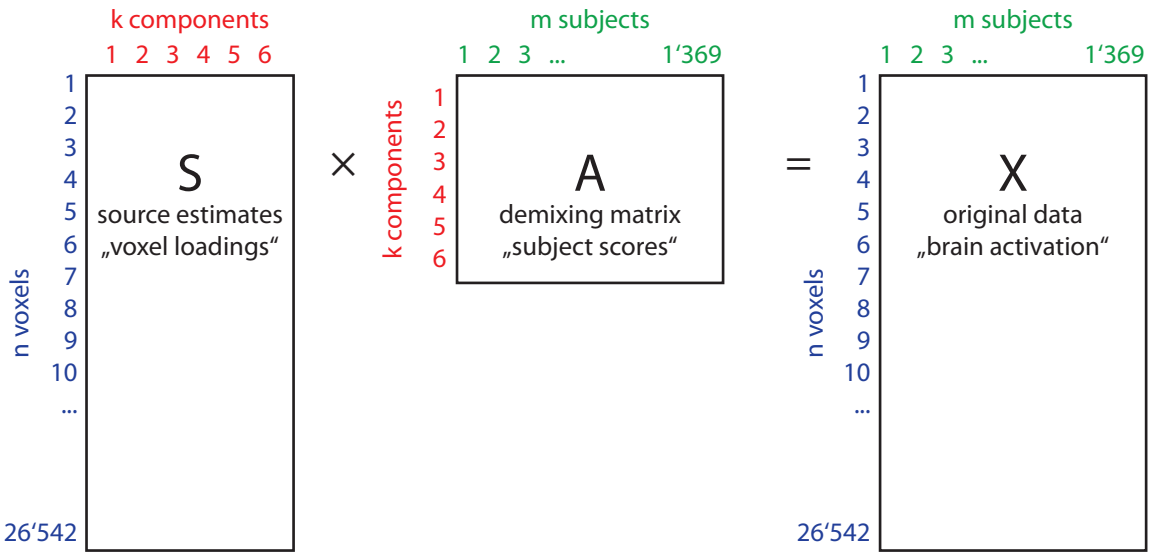

Figure 6. The estimates that result from an ICA decomposition of brain activation data in from 1'369 subjects across 26'542 voxels.

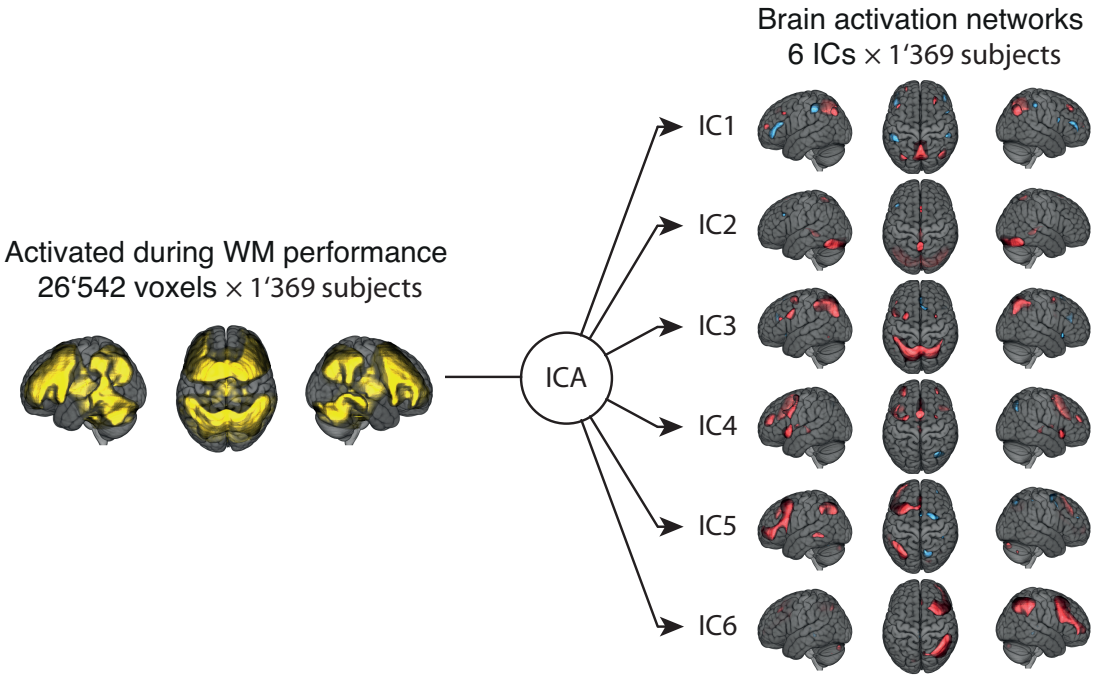

Figure 7. ICA decomposition of brain activation during a working memory task resulted in voxel loadings for each IC. The three-dimensional visualizations of brain imaging data in a semi-transparent brain were created using the software MRIcroGL.

4. Original Research Papers

4.1 Distinct working memory brain networks in healthy young adults

Egli, T., Coynel, D., Spalek, K., Fastenrath, M., Freytag, V., Heck, A., Loos, E., Auschra, B., Papassotiropoulos, A., de Quervain, D. J.-F., & Milnik, A. (2018). *eNeuro, 5*(1), e0222-17.2018

Cognition and Behavior

Identification of Two Distinct Working Memory-Related Brain Networks in Healthy Young Adults

Tobias Egli,[1,2] David Coynel,[2,3] Klara Spalek,[2,3] Matthias Fastenrath,[2,3] Virginie Freytag,[1,2] Angela Heck,[1,2,4] Eva Loos,[2,3] Bianca Auschra,[1,2] Andreas Papassotiropoulos,[1,2,4,5] Dominique J.-F. de Quervain,[2,3,4] and Annette Milnik[1,2,4]

DOI:http://dx.doi.org/10.1523/ENEURO.0222-17.2018

[1]Division of Molecular Neuroscience, Department of Psychology, University of Basel, CH-4055 Basel, Switzerland, [2]Transfaculty Research Platform Molecular and Cognitive Neurosciences, University of Basel, CH-4055 Basel, Switzerland, [3]Division of Cognitive Neuroscience, Department of Psychology, University of Basel, CH-4055 Basel, Switzerland, [4]Psychiatric University Clinics, University of Basel, CH-4055 Basel, Switzerland, and [5]Department Biozentrum, Life Sciences Training Facility, University of Basel, CH-4056 Basel, Switzerland

Abstract

Working memory (WM) is an important cognitive domain for everyday life functioning and is often disturbed in neuropsychiatric disorders. Functional magnetic resonance imaging (fMRI) studies in humans show that distributed brain areas typically described as fronto-parietal regions are implicated in WM tasks. Based on data from a large sample of healthy young adults ($N = 1369$), we applied independent component analysis (ICA) to the WM-fMRI signal and identified two distinct networks that were relevant for differences in individual WM task performance. A parietally-centered network was particularly relevant for individual differences in task measures related to WM performance ("WM dependent") and a frontally-centered network was relevant for differences in attention-dependent task performance. Importantly, frontal areas that are typically considered as key regions for WM were either involved in both WM-dependent and attention-dependent performance, or in attention-dependent performance only. The networks identified here are provided as publicly available datasets. These networks can be applied in future studies to derive a low-dimensional representation of the overall WM brain activation.

Key words: cognition; functional networks; ICA; n-back; working memory

Significance Statement

Fronto-parietal brain regions are typically involved when performing working memory (WM) related tasks. Within these fronto-parietal brain regions we have identified two networks that show distinct functional characteristics. Whereas frontal areas are often considered as key regions for WM, we show that frontal areas were either involved in both WM-dependent and attention-related performances or in attention-related performance only. A predominately parietally-centered network was the key region for WM-dependent performance. Due to the large sample size of $N = 1369$ healthy young adults, we can provide robust estimates of these networks which can be applied in future studies.

Introduction

Working memory (WM) describes the ability to temporarily maintain and manipulate a limited amount of information (Baddeley, 2012; Eriksson et al., 2015). It comprises a mental representation of our current environment that can be integrated with previous experiences. Impaired WM leads

Received June 27, 2017; accepted January 18, 2018; First published January 30, 2018.
The authors declare no competing financial interests.

Author contributions: D.Q. and A.P. designed research; K.S. and E.L. performed research; T.E., A.M., D.C., M.F., V.F., A.H., and B.A. analyzed data; T.E., D.C., K.S., M.F., V.F., A.H., E.L., B.A., A.P., D.J.-F.d.Q., and A.M. wrote the paper.

to deterioration in everyday life functioning. Correspondingly WM is affected in neuropsychiatric disorders such as schizophrenia (Lee and Park, 2005; Forbes et al., 2009; Mesholam-Gately et al., 2009; Van Snellenberg et al., 2016), depression (Marazziti et al., 2010), and attention-deficit hyperactivity disorder (Alderson et al., 2013). Furthermore, white matter microstructure is associated with WM performance and activity in WM related regions (Charlton et al., 2010; Vestergaard et al., 2011; Darki and Klingberg, 2015). In contrast, impairment of white matter integrity comes along with a decrease in WM performance and alterations in the activity of WM-related brain regions (Palacios et al., 2012).

Functional magnetic resonance imaging (fMRI) experiments show that WM-related tasks robustly activate the lateral and medial premotor cortex, dorsolateral prefrontal cortex (DLPFC) and ventrolateral PFC, frontal pole, as well as medial and lateral posterior parietal cortex (Owen et al., 2005; Wager and Smith, 2003; Rottschy et al., 2012). This broad WM network (WMN) of activated brain regions has been studied extensively, including the use of meta-analytical approaches (Yarkoni et al., 2011; Rottschy et al., 2012). Several studies have observed associations of WM performance with mainly parietal or fronto-parietal brain activation (Klingberg et al., 2002; Todd and Marois, 2004, 2005; Nagel et al., 2005; Palacios et al., 2012; Satterthwaite et al., 2013; Zou et al., 2013; Ullman et al., 2014; Darki and Klingberg, 2015; Huang et al., 2016) in children as well as in adults. Recent studies suggest that frontal and parietal regions differ regarding their contributions to WM. Neuronal recordings in the PFC and the lateral intraparietal (LIP) region of monkeys showed that encoded stimuli were retained in both regions, with more task-specific mnemonic encoding in the LIP as compared to the PFC (Sarma et al., 2016). Another study provided causal evidence for differing roles of parietal and frontal regions in attentional aspects of WM processing, by applying transcranial direct current stimulation. Stimulating the right parietal cortex increased the amount of information maintained in the visual WM, whereas stimulating the right PFC improved focusing on relevant information and directing attention away from irrelevant stimuli (Li et al., 2017). In addition, measuring the directed connectivity between the DLPFC and superior parietal lobule (SPL) during a visual WM task hinted toward a top-down drive from DLPFC to SPL that increased with WM load (Kundu et al., 2015). These insights were based on a priori defined regions of interest (ROIs) and therefore described functional properties of separate brain regions.

Importantly, the human brain is organized in functional intrinsic networks that are relatively stable during resting state as well as task execution (Cole et al., 2014; Cole et al., 2016), can exhibit spatial overlaps (Yeo et al., 2014), and are also affected by neurodegenerative diseases (Seeley et al., 2009; Zhou et al., 2012). Hence, instead of applying a ROI-based approach, we used independent component analysis (ICA) to identify distinct networks within the WMN, as measured by the n-back task, based on data from a large ($N = 1369$) sample of healthy young adults. ICA decomposition is a data-driven unbiased approach to retrieve a low-dimensional representation of a dataset, resulting in statistically independent signals (Kong et al., 2008). We included both cortical and subcortical regions into the ICA decomposition to retrieve maximally unbiased estimates of brain networks. To functionally classify these networks, we used cognitive performance measurements of our subjects. We verified the stability of our results using bootstrapping and cross-validation procedures. Furthermore, we assessed whether microstructural differences of white matter, measured by diffusion tensor imaging (DTI), were associated with activation differences in the estimated networks. Finally, we compared the networks estimated in our study with results from an extensive meta-analysis of neuroimaging studies on WM brain activation (Rottschy et al., 2012) and with networks derived from NeuroSynth, a meta-analytical platform comprising a large variety of different fMRI studies (Yarkoni et al., 2011). All results obtained (univariate statistics and estimates from the ICAs) are available as parametric maps stored on NeuroVault (http://neurovault.org/collections/EYCSLZUZ/; Gorgolewski et al., 2016) and can be used for future studies. The WMN-IC estimates can be used to derive a low-dimensional representation of the overall WM brain activation.

Materials and Methods

Study and sample description

We used data from a single-center fMRI study that aims to identify biological correlates of cognitive performance by combining imaging data with genetics data; note that no genetic data were used here. With respect to the cognitive performance measurements, this study emphasizes on WM and episodic memory performance. The sample consisted of healthy young adults from the general population. We analyzed data of 1369 subjects (mean age: 22.4, range: 18–35; 841 females; the experiment took place at the University Hospital of Basel) after excluding subjects with incomplete behavioral data ($N = 28$), with cognitive measurements (WM, attention, reaction time, episodic memory, recognition memory) lying 4 SDs above or below the average ($N = 15$), with corrupted imaging data ($N = 38$, see below, fMRI preprocessing and first-level analyses of the n-back task), or with incomplete imaging data ($N = 6$, see below, fMRI preprocessing and first-level analyses of the n-back task). Subjects were free

This work was supported by grants from the Swiss National Science Foundation and the University of Basel.

Acknowledgments: Calculations were in part performed at sciCORE scientific computing core facility at University of Basel (http://scicore.unibas.ch/). We thank Elmar Merkle, Christoph Stippich, and Oliver Bieri for granting access to the fMRI facilities of the University Hospital Basel.

Correspondence should be addressed to either of the following: Tobias Egli, Division of Molecular Neuroscience, University of Basel, Birmannsgasse 8, CH-4055 Basel, Switzerland, E-mail: tobias.egli@unibas.ch; or Annette Milnik, Division of Molecular Neuroscience, University of Basel, Birmannsgasse 8, CH-4055 Basel, Switzerland, E-mail: annette.milnik@unibas.ch.

DOI:http://dx.doi.org/10.1523/ENEURO.0222-17.2018

from any neurologic or psychiatric illness and did not take any medication (except oral contraception) at the time of the experiment. Women using hormonal contraceptives (e.g., oral, spiral, patch) and naturally cycling women were included in the study without restrictions. The ethics committees of the Cantons of Basel-Stadt and Basel-Landschaft approved the study. Advertising for study participation was conducted mainly in the University of Basel. Written informed consent was obtained from all subjects before participation.

Experimental procedure

After receiving general information about the study and giving their written informed consent, participants were first instructed and then trained on a picture-rating task and an n-back task. This training was done outside of the MR scanner. After training, participants were positioned in the scanner. All subjects wore earplugs and headphones during MR scans to reduce scanner noise. The participants were instructed not to move during the scans. Small foam pads were used for additional head fixation. We used MR-compatible LCD goggles (VisualSystem, NordicNeuroLab) to present the behavioral tasks inside the scanner. Vision correction was used if necessary. The participants first performed the picture-encoding task in which they had to rate pictures. Afterward they performed the WM task (n-back). During this first fMRI session participants spent a total of 30 min in the scanner (20 min on the picture-rating task, 10 min on the n-back task). Participants then left the scanner and performed an unannounced free recall task of the previously presented pictures (without any time restriction). On finishing the free recall, subjects were instructed and trained on a picture recognition task. This training was done outside of the scanner. Subjects were then positioned in the MR scanner a second time. The picture recognition task lasted 20 min and was followed by T1 (anatomic MRI) and DTI measurements for a further 20 min. The total length of the experimental procedure ranged from 3 to 4.5 h per subject. Participants were rewarded with 25 Swiss Francs per hour for participating.

WM task description

We used two different conditions of a verbal n-back task. The 0-back condition required participants to respond to the occurrence of the letter "x" as target stimulus (both lower- and uppercase) in a sequence of letters (e.g., N – p – X – g. . .); all other letters were nontarget stimuli. In the 2-back condition subjects had to indicate whether the current letter and the letter presented two places prior in the sequence were identical (target stimulus) or not (nontarget stimulus); e.g., S – f – s – g. . . Each condition was measured in six blocks. Every block consisted of 14 stimuli. In each block, three target stimuli and 11 nontarget stimuli were presented (quasi)-randomly; the frequency of lure trials (i.e., the most recent letter matches the letter one or three positions back) was set to 17.9% (15 out of 84 stimuli) in the 2-back condition. Each block started with an instruction of 5 s and had a total duration of 33 s. Each stimulus was presented for 500 ms with a 1500-ms interstimulus interval showing a black screen. The sequence of 2-back and 0-back blocks was randomized and a break of 20 s was added after every second block. The subjects used a button-box to indicate each stimulus either as "target" or as "nontarget." The data were disregarded if responses were missing (1) in >30% of all stimuli across all twelve blocks of the task, (2) in >30% of target stimuli in at least three blocks, or (3) in >30% of nontarget stimuli in at least three blocks. Task performances were defined as D-prime measures (Macmillan and Creelman, 1990). These measures account for false alarms and were calculated separately for the 0-back and 2-back conditions. The task performance ranged from -0.34 to 4.34 ($M = 2.53$; $Md = 2.47$) for the D-prime 2-back and from 1.56-4.34 ($M = 3.65$; $Md = 3.76$) for the D-prime 0-back. We also used the difference in performances between D-prime 2-back and D-prime 0-back, which ranged from -4.10 to 1.38 ($M = -1.13$, $Md = -1.10$). As a measure of difference in reaction times, we used the subtracted reaction time of the two conditions (reaction time 2-back – 0-back), which varied from -37.26 to 602.32 ms ($M = 126.15$ ms; $Md = 104.25$ ms).

Descriptions of picture-related tasks

The picture-rating task required the participants to rate 72 pictures of positive, neutral, and negative valence (24 per valence group). While watching the pictures the participants rated each picture's emotional valence (positive, neutral, negative) and the perceived arousal (low, middle, high) on separate three-point Likert scales. Approximately 10 min later, the subjects were instructed to describe as many of these pictures as possible and in as much detail as possible by using keywords or short sentences (free recall of pictures). Based on these descriptions two independent and blinded raters identified the number of correctly recalled pictures (Cronbachs α between the two raters was 0.91 to 0.98). A third independent rater decided on ambiguously scored pictures. The number of correctly recalled pictures served as a measure of episodic memory performance (range: 5–55 pictures; $M = 30.77$; $Md = 31$). This free recall of the pictures was conducted in several different rooms; the effect of the different rooms on the free recall performance was regressed out before running the analyses.

In the picture recognition task, 144 pictures in total were presented: the 72 previously seen pictures and 72 new pictures. The participants rated these pictures as remembered, familiar, or new on a three-point Likert scale. Item familiarity corresponds to the number of previously seen pictures that were identified as "familiar," corrected for the number of new pictures that were wrongly rated as familiar. The item familiarity performance ranged from -32 to 48 ($M = 3.53$; $Md = 2$). Both, the episodic memory task and the familiarity memory task used photographic pictures of positive, neutral, and negative valence selected from the International Affective Picture System (IAPS; Lang et al., 2008). In-house standardized pictures additionally complemented the neutral picture set to equate the stimuli for visual complexity and content (e.g., human presence).

Description of further task performances and covariates

A total of 90.3% of the participants used their right hand while performing the tasks in the scanner, 9.7% used their left hand. The self-reported body mass index (BMI) ranged from 16.6 to 36.3 ($M = 22.19$; $Md = 21.80$). We assessed distinct chronotypes on a two-point Likert scale: subjects classified themselves either as "eveningness" (69.8%) or as "morningness" (30.2%) chronotype. The self-reported sleep duration ranged from 3.75 to 12 h ($M = 7.96$; $Md = 8$). Self-reported smoking was measured on a five-point Likert scale ranging from 1 (never) up to 5 (20 cigarettes per day); the relative frequencies per category were: (1) 65%, (2) 23%, (3) 5.2%, (4) 6.8%, (5) 0.7%. After finishing all tasks, the perceived overall task difficulty and the overall motivation of the subjects were measured on five-point Likert scales ranging from 1 (not at all) up to 5 (very). The relative frequencies per category for task difficulty were: (1) 9.2%, (2) 40.3%, (3) 38.1%, (4) 12%, (5) 0.3% and for motivation were: (1) 0%, (2) 0.5%, (3) 6.6%, (4) 44.4%, (5) 48.4%.

(f)MRI data acquisition

All functional and structural images were acquired on the same Siemens Magnetom Verio 3 T whole-body MR unit (12-channel head coil). Blood oxygen level-dependent fMRI was acquired using a single-shot echoplanar sequence along with generalized auto-calibrating partially parallel acquisition (GRAPPA), using the following parameters: echo time (TE) = 25 ms, field of view (FOV) = 22 cm, acquisition matrix = 80 × 80 (interpolated to 128 × 128, voxel size 2.75 × 2.75 × 4 mm^3) and with an acceleration factor of 2. We used an ascending interleaved sequence with repetition time (TR) = 3000 ms ($\alpha = 82°$) measuring 32 contiguous axial slices that were placed along the anterior-posterior commissure plane based on a midsagittal scout image. A magnetization-prepared rapid acquisition gradient echo T1-weighted image was acquired using the following parameters: TR = 2000 ms, TE = 3.37 ms, TI = 1000 ms, flip angle = 8°, 176 slices, FOV 256 mm, and voxel size = 1 mm^3. Automatic segmentations of cortical and subcortical structures were obtained using FreeSurfer 4.5 (v4.5, http://surfer.nmr.mgh.harvard-.edu/; RRID:SCR_001847; Fischl, 2012), and labeling was based on the Desikan Atlas (Desikan et al., 2006).

fMRI preprocessing and first-level analyses of the n-back task

After visual inspection by three raters, 38 participants were excluded due to corrupted T1-weighted images (movement or anatomic abnormalities). MR images were preprocessed with SPM8 (Statistical Parametric Mapping, Wellcome Trust Center for Neuroimaging; http://www.fil.ion.ucl.ac.uk/spm/) implemented in MATLAB R2011b (MathWorks). Slice-time correction to the first slice and realignment were applied using the "register to mean" option. Coregistration of the averaged realigned time series to the structural image ensured spatial alignment of functional and structural images. Subject-to-template normalization was done using DARTEL (Ashburner, 2007), which allows registration to both cortical and subcortical regions and has been shown to perform well in volume-based alignment (Klein et al., 2009). Normalization incorporated the following four steps. (1) Structural images of each subject were segmented using the "New Segment" procedure in SPM8. (2) The resulting gray and white matter images were used to derive a study-specific group template. The template was computed from a subgroup of 1000 subjects (Heck et al., 2014), which were part of the 1369 subjects in the present study. (3) An affine transformation was applied to map the group template to MNI space. (4) Subject-to-template and template-to-MNI transformations were combined to map the functional images to MNI space. The functional images were smoothed with an isotropic 8 mm full-width at half-maximum (FWHM) Gaussian filter. Intrinsic autocorrelations were accounted for by AR(1) and low-frequency drifts were removed via high-pass filter (time constant 128 s). Separate regressors were constructed for the 0- and 2-back conditions comprising a boxcar reference wave form convolved with a canonical hemodynamic response function. Events during the presentation of the instruction as well as movement regressors from spatial realignment were modeled separately. To measure WM-related brain activation we calculated the difference between the 2-back and 0-back parameter estimates for each subject and voxel (first-level 2-back – 0-back contrast). Performance measurements were not included in the first-level analyses.

fMRI group-level analysis

All further analyses were conducted using the statistical software environment R (3.2.2; RRID:SCR_001905). The 2-back – 0-back contrast parameters from the first-level analyses of $N = 1375$ subjects and of $N = 71222$ voxels entered the group analyses. Data of six subjects were removed from the analyses because of high numbers of missing voxels (>4 SD above average). For the remaining $N = 1369$ subjects, we then restricted all analyses to voxels without missing values ($N = 55{,}614$ voxels). Based on one-sample t tests, we identified all voxels that were more active in the 2-back in comparison to the 0-back condition when applying FDR correction ($\alpha = 5\%$).

Across the timespan of the data acquisition, the gradient coils were changed twice (hardware batches), and parts of the scanner's software configuration were changed once (software batches). Additionally, the scanner console displayed irregularities during the data acquisition in a small group of subjects (processing batches). We regressed out these potential group-effects from the voxel-signal; we used the standardized residuals to perform the ICA decomposition and the association analyses.

Identification of distinct WMN subnetworks by using ICA decomposition

We investigated the distribution of 2-back – 0-back contrast parameter estimates by measuring the skewness, kurtosis, and Shapiro–Wilk tests. The data were highly skewed across subjects (−2.59–2.34) and voxel (−2.03–2.84), showed a high kurtosis across subjects (2.98–49.11), and voxel (3.26–23.06) and deviated con-

siderably from normal distribution across subjects (Shapiro–Wilk test: range W, 0.85–1.00; range $-\log_{10}(p)$, 0.33–33.72) and voxel (Shapiro–Wilk test: range W, 0.80–1.00, range $-\log_{10}(p)$, 6.31–60.60). Because of the strong non-Gaussian components of the 2-back – 0-back contrast parameters, we used ICA as dimensionality reduction method. Applied to a matrix X of m observations (subjects) and n variables (voxels), ICA estimates a matrix of $k \times n$ latent sources S that underlie the variables, holding the source estimates (referred to as voxel loadings throughout the paper) as statistically independent from each other as possible (Engreitz et al., 2010). In addition to the source estimates, ICA also yields a matrix of $m \times k$ mixing coefficients A (referred to as subjects scores throughout the paper) for each IC. The mixing coefficients of a particular component depict the projection of the original data onto this component's estimated source, such that $X = AS$ (Hyvärinen and Oja, 2000). By applying ICA decomposition to a matrix of 2-back – 0-back contrast estimates, containing rows of voxels and columns of subjects, our source estimates (voxel loadings) described statistically independent latent sources that underlie the contrast estimates. Accordingly, each component's mixing coefficients described the activity strength of each component for each subject (Chiappetta et al., 2004). Subjects with high-contrast estimates in the voxels that load highly onto a particular IC in the positive direction obtained elevated scores for this IC. Hence, we interpreted the subject scores as a measure of coactivation in the voxels that loaded onto the IC.

We first applied PCA to determine the number of components to be extracted by the ICA. After visually inspecting the scree plot of the Eigenvalues we decided to retrieve six components. We performed ICA to retrieve these six ICs using the fastICA algorithm (R-package "fastICA"; Hyvärinen and Oja, 2000) with centering and scaling of the variables as well as applying a PCA and whitening of the data. Since the direction of ICA estimates is arbitrary, we recoded all estimated ICs with the result that the voxels with the highest absolute loadings displayed positive loadings. We retained the source estimates ("voxel loadings") and mixing coefficients ("subject scores") of the extracted ICs (WMN-ICs) for further analyses. Accordingly, every voxel exhibited a voxel loading for each of the six WMN-ICs. For visualization purpose and for anatomic annotation, we determined the voxel loadings with the 10% most extreme absolute values ($|z| > 1.47$), when considering all six ICs. All association analyses were conducted on unthresholded WMN-ICs.

Cortical and subcortical labeling of the WMN-ICs

Labeling of gray matter brain regions was based on a population-averaged probabilistic atlas. The atlas comprises a total of $N = 87$ distinct cortical and subcortical brain regions from both hemispheres. Each of the $N = 55{,}614$ voxels was assigned to one of these anatomic brain regions. Voxels for which the probability to belong to a given brain region was below 25% ($N = 2926$) or that were not located within cortical or subcortical regions ($N = 21{,}451$) were excluded, resulting in $N = 31{,}237$ voxels used for anatomic labeling. For each WMN-IC, we grouped voxels that showed the 10% most extreme values (see above) into clusters of adjacent voxels (WMN-IC clusters). Within each WMN-IC cluster and for each anatomic brain region, we determined the absolute number of voxels that belonged to this cluster and were annotated with this region. We report only brain regions comprising >10 voxels of a WMN-IC cluster. We also calculated the percentage of voxels per WMN-IC cluster and anatomic brain region by dividing the absolute number of voxels by the total number of voxels labeled with the anatomic brain region across the $N = 31{,}237$ voxels.

The used population-average probabilistic anatomic atlas was built by automatic gray matter segmentation of the subjects' T1-weighted images. Each participant's T1-weighted image was first automatically segmented into cortical and subcortical structures using FreeSurfer (v4.5, http://surfer.nmr.mgh.harvard.edu/; RRID:SCR_001847; Fischl, 2012). Labeling of the cortical gyri was based on the Desikan–Killiany Atlas (Desikan et al., 2006), yielding 35 regions per hemisphere. We also labeled 17 subcortical regions, following Fischl et al., (2002). The segmented T1 image was then normalized to the study-specific anatomic template space using the subject's previously computed warp field, and affine-registered to the MNI (Montreal Neurologic Institute) space (see above, fMRI preprocessing and first-level analyses of the n-back task). Nearest-neighbor interpolation was applied, to preserve labeling of the different structures. The normalized segmentations were finally averaged across subjects, to create a population-average probabilistic atlas. Each voxel of the template could consequently be assigned a probability of belonging to a given anatomic gray matter-segmented structure, based on the information of $N = 1000$ subjects that are part of the samples included in this study.

Association with task performance measures

We assessed the associations of each WMN-IC with performance measurements of multiple behavioral tasks and several covariates using a multiple linear regression model for each WMN-IC. For each WMN-IC, the scores per subject were used as the dependent variables. The task performance measurements and covariates were assigned as independent variables. To reduce multicollinearity between the independent variables and covariates, we excluded strongly correlated variables ($|r_{Pearson}| > 0.5$).

The following behavioral task performances were included: (1) n-back performances (D-prime 2-back; D-prime 0-back), (2) n-back reaction time (difference between reaction times during 2-back condition and 0-back condition), (3) episodic memory, and (4) item familiarity. We first calculated linear models with the difference in 2-back and 0-back performances as a single predictor. To estimate the associations with 2-back and 0-back performances individually, we also included both performance measurements separately in the model. We further included the following covariates in the analyses: (5) Sex, (6) age at the time of investigation, (7) hand used for task performance, (8) motivation, (9) perceived task difficulty, (10) smoking behavior, (11) usual sleep duration, (12) chronotype, and

(13) BMI. Since the scores of the WMN-ICs were correlated ($r^2 < 0.11$; see results section "Identification of distinct WM-task networks"), we additionally included the scores of the five remaining ICs as covariates in all analyses. The regression models thus comprised 18 predictors when including the difference in 2-back and 0-back performances as a single predictor, and 19 predictors when including separate predictors for 2-back and 0-back performances.

To retrieve standardized regression coefficients (subsequently referred to as regression coefficient or β), all variables were z-transformed. By including all predictors and covariates in one linear model, we estimated the association between each variable and the WMN-IC while keeping all other included variables constant. Testing of significance for the behavioral task performances was conducted using *t* tests. We report FDR-corrected *p* values for associations of WMN-ICs with task performance ($\alpha < 5\%$, correcting for 108 tests based on 18 predictors $\times$ six WMN-ICs with the difference in 2-back and 0-back performances as a single predictor; correcting for 114 tests based on 19 predictors $\times$ six WMN-ICs with 2-back and 0-back performances as separate predictors).

We used the same linear models, but without including the WMN-IC scores as covariates, to estimate the univariate association of each voxel with D-prime 2-back and D-prime 0-back performances. We applied FDR correction ($\alpha = 5\%$) to account for 371588 independent statistical tests, based on 14 predictors $\times$ 26542 voxels.

ICA bootstrapping

We assessed the stability of the WMN-ICs and of their associations with behavioral measures using a bootstrapping approach. We repeated the following procedure 100 times for two different sizes of the subsamples $N_{subsample} = [100, 684]$. We (1) randomly divided the sample into two subsamples of sizes $N_{subsample}$ (sampling without replacement, no intersection between the subsamples); (2) for both subsamples, we estimated six ICs; and (3) calculated linear models of the IC estimates against behavioral measures and covariates as described above; and (4) for each IC of both subsamples, we identified the best-matching IC of the total sample. We correlated the source estimates (i.e., voxel-loadings) of these matched ICs from the two subsamples.

ICA cross-validation

We projected the information from WMN-ICs that were estimated across $N = 1269$ subjects onto smaller groups of $N = 100$ subjects. We repeated the following procedure 100 times: in each run, (1) we randomly divided the sample into the larger and the smaller subsamples; (2) we estimated six ICs from the WMN in the larger subsample; (3) the ICA estimates were then projected onto the 2-back – 0-back contrast estimates of the smaller subsample; and (4) the resulting projected scores of WMN-IC3 and WMN-IC4 were then regressed against behavioral task performances (D-prime 2-back, D-prime 0-back) and covariates (sex, age) in the smaller subsample. This yielded the percentage of runs in which the projected scores showed significant ($p_{nominal} < 0.05$) associations with task performance measures. To retrieve empirical *p* values for the cross-validation, we repeated the 100 cross-validations 1000 times, after permutation of the task performance measurements. We used the percentages of associations between projected scores and permuted performance measurements as a null distribution.

Association of the WMN-ICs with white matter microstructure

Diffusion volumes were acquired for a subset of $N = 657$ subjects using a single-shot EPI sequence, and consisted of 64 diffusion-weighted volumes with b = 900 s/mm^2 and one unweighted volume (b = 0). We used the following acquisition parameters: TR = 9 s; TE = 82 ms; FOV = 320 mm; GRAPPA R = 2.0; voxel size = 2.5 $\times$ 2.5 $\times$ 2.5 mm^3. Two participants were excluded due to excessive movement during the DTI acquisition. Diffusion-weighted images were analyzed using FSL (4.1.7; RRID: SCR_002823; Jenkinson et al., 2012). Images were coregistered to the reference unweighted volume (b = 0) using an affine transformation for correction of head motion and eddy current induced image distortion. Maps of fractional anisotropy (FA) were obtained from the diffusion tensor model for further analyses. FA is an estimate of the directional dependence of diffusion (Basser, 1995). It reflects aspects of white matter microstructure that are related to fiber orientation (Jones et al., 2013) and can be modulated by myelination (Beaulieu, 2002). We obtained 70 cortical white matter-segmented regions (35 regions per hemisphere) from the FreeSurfer v4.5 *wmparc* files. Anatomic labels for the white matter segmentations corresponded to the labels of gray matter segmentations adjacent to the corresponding white matter segmentation. We used the averaged FA values per region for the following analyses. Sixteen participants were excluded due to missing FA measures in any of the white matter-segmented brain regions. Complete datasets (behavior and imaging) were available for $N = 614$ participants. For each white matter-segmented brain region and each WMN-IC, we calculated linear regression models with the WMN-IC's scores per subject as dependent variables and the FA estimate as independent variable. Sex, age, handedness, intracranial volume, and scores of the remaining WMN-ICs were used as covariates. We tested separately for each WMN-IC whether the *p* values of the associations between the 70 FA values and WMN-IC scores deviate from the uniform distribution that is expected for continuous data under a simple null hypothesis (Murdoch et al., 2008). The resulting *p* values were FDR corrected for six tests ($\alpha < 0.05$). We additionally calculated empirical *p* values based on the number of nominally significant associations for each WMN-IC after permuting the WMN-IC scores 10000 times, applying FDR correction ($\alpha < 0.05$) for six independent tests.

Description and analysis of the NeuroSynth database

NeuroSynth is a publicly available database currently comprising data from 11406 fMRI studies summarized in 3107 fMRI meta-analyses for commonly used terms (RRID:SCR_006798; Yarkoni et al., 2011). We obtained the NeuroSynth data files (database.txt; features.txt; ver-

sion 0.6, released July, 2015) as well as the reverse inference maps of all 3107 meta-analyses. The reverse inference maps of the meta-analyses describe for each voxel the probability of the term being used in the available studies given the activations in the voxel across the studies; these inference maps contain estimates for voxels showing FDR-corrected ($\alpha = 0.01$) significant associations. We first selected all terms that were reported in at least 250 studies at a high frequency (>1 in 1000 words). For these terms we filtered for all reverse inference maps that comprise at least 1200 FDR-corrected significant voxels (out of 228,453 voxels, > 0.5%; voxel size 2 × 2 × 2 mm). After applying these filter-steps we used the meta-analytic results of 233 terms for the further analyses. We applied z-transformation to the probability estimates for each term before applying PCA. After visually inspecting the scree plot of the PCA (see results section "Comparison of the WM-task networks with external datasets"), we decided to extract 16 components. After whitening of the data we applied ICA decomposition on the probability estimates using the fastICA algorithm (R-package fastICA; Hyvärinen and Oja, 2000) to retrieve 16 networks that were based on the results of the 233 meta-analyses. Since the direction of ICA estimates is arbitrary, we recoded all estimated ICs with the result that the voxels with the highest absolute loadings displayed positive loadings. The mixing coefficients (score per term) were used to characterize each component (NeuroSynth IC-topic).

The uncorrelated and statistically independent source estimates (loadings per voxel) were coregistered to the image space of our functional MRI data by applying affine transformation with NiftyReg (http://cmictig.cs.ucl.ac.uk/wiki/index.php/NiftyReg; RRID:SCR_006593; Modat et al., 2010). We tested the overlap between the 16 NeuroSynth networks and the WMN derived from our functional MRI data by calculating the percentage of voxels that show high loadings on the NeuroSynth networks ($|z| > 0.70$; i.e., the 10% most extreme absolute values across all NeuroSynth ICs; the same threshold was used to visualize the NeuroSynth ICs) and were additionally located in the WMN. Furthermore, we compared the loadings per voxel between the NeuroSynth networks and the WMN-ICs (shared variance r^2). We retrieved subject-wise scores for the NeuroSynth IC-topics in our study sample by projecting the NeuroSynth ICA estimates onto the 2-back – 0-back contrast parameter estimates of our subjects. The projected scores for the NeuroSynth IC-topic were regressed against the subjects' task performance measures using multiple linear regression models (including sex, age, hand used for the task, motivation, perceived task difficulty, smoking behavior, usual sleep duration, chronotype and BMI as covariates). The resulting p values were FDR corrected ($\alpha = 0.05$) for 224 independent tests, based on 14 predictors × 16 NeuroSynth ICs.

Brain images

Figures of clustered voxels within a semitransparent brain (MNI 152 template) were produced using MRIcroGL (http://www.mccauslandcenter.sc.edu/mricrogl/; RRID:SCR_002403) after smoothing (3 mm smoothing kernel) using the R-packages "fslr" (Muschelli et al., 2015) and "oro.nifti" (Whitcher et al., 2011). All brain images are displayed within the MNI152 template and according to neurologic convention (left hemisphere displayed on the left side).

Data repository

Parametric maps of the main findings (group-activation t values for the 2-back – 0-back contrast parameter; β values for associations between the 2-back – 0-back contrast parameters and the 2-back as well as the 0-back performances; z values describing voxel loadings of the ICs) are stored online in the public repository NeuroVault (RRID:SCR_003806; Gorgolewski et al., 2016) and can be retrieved for use in future studies (http://neurovault.org/collections/EYCSLZUZ/).

Results

We used two different conditions of a verbal n-back task. The 0-back condition required participants to respond to the occurrence of the letter x (both lower- and uppercase) in a sequence of letters (e.g., N – p – X – g. . .). This control condition requires very low WM load and was used as a measure of attention. In the 2-back condition subjects had to indicate whether the currently presented letter and the letter two places prior in the sequence were identical or not (e.g., S – f – s – g. . .). This condition requires online monitoring, updating, and manipulation of remembered information and is therefore assumed to involve key WM-related processes (Owen et al., 2005). Task performances were defined as D-prime measures (Macmillan and Creelman, 1990) that account for false alarms, calculated separately for the 0-back and 2-back conditions. Both behavioral measurements were correlated with a medium effect size ($r_{Pearson} = 0.35$; 12% shared variance). The 0-back performance is also referred to as "attention-related" and the 2-back performance is also referred to as "WM-related" task performance in the following sections.

fMRI group-level analysis of the WM-task activation

The fMRI analyses were based on the 2-back – 0-back contrast parameter estimates. We first applied voxel-wise ($N = 55614$ voxels) one-sample t tests to the contrast parameter estimates. Here, due to the large sample size ($N = 1369$), the whole-brain signal was virtually separated into voxels that were more active in the 2-back condition, and voxels that were more active in the 0-back condition (see "t value contrast 2-back – 0-back" in NeuroVault). The WMN is typically defined as voxels that are more active in the 2-back condition in comparison to the 0-back condition (Rottschy et al., 2012); the 0-back condition is included to control for sensory-motor processes and attention (Miller et al., 2009). The WMN identified with our data were defined as the 2-back positive voxels of the 2-back – 0-back contrast parameter estimates (whole brain FDR-corrected $\alpha < 5\%$; $N = 26,542$ voxels; Fig. 1*A*). This WMN comprised most of the FDR-corrected meta-analytic result for the term "working memory" acquired from NeuroSynth (Yarkoni et al., 2011; Fig. 1*B*): 98% of the WMN voxels derived from NeuroSynth were located within the WMN obtained from our data.

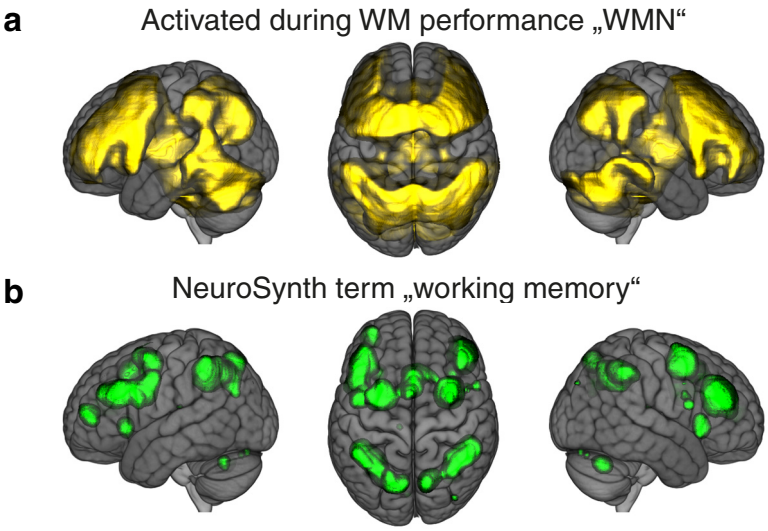

Figure 1. WMNs. ***A***, Brain regions that were more strongly activated during the 2-back condition in comparison to the 0-back condition in our sample (2-back – 0-back contrast one-sample *t* tests FDR corrected, $\alpha = 0.05$). ***B***, Meta-analytic results for the term working memory retrieved from NeuroSynth (reverse inference, FDR corrected, $\alpha = 0.01$). The brain images are displayed within the MNI152 template and according to neurologic convention.

Identification of distinct WM-task networks

To identify separable networks of brain activation within the WMN we applied ICA as a dimensionality reduction method. ICA decomposition is a data-driven unbiased approach that models observations as a linear combination of latent components (Engreitz et al., 2010), which are as statistically independent and uncorrelated as possible (Hyvärinen and Oja, 2000). We applied ICA onto the 2-back – 0-back contrast estimates of our subjects. Each voxel obtained one loading per IC, and each subject obtained one score per IC. The ICs were statistically independent and uncorrelated with regard to their voxel loadings. Accordingly, a voxel's loading in a particular IC did not yield any information regarding this voxel's loading in any other IC. When illustrating the voxel loadings of the ICs, we concentrated on the voxels with the most extreme 10% of loadings. Whenever a subject showed increased activation in the brain regions that loaded highly onto an IC in the positive direction, the subject received an elevated positive score for the specific IC. Accordingly, the subject scores of an IC represented a measure of coactivation across the voxels that loaded onto this IC. We therefore interpreted the estimated ICs as networks of coactivated brain regions.

Whitening of the data was done based on a principal component analysis (PCA) before applying the ICA. After visually inspecting the Eigenvalues of the PCA (Fig. 2*A*) we decided to extract six ICs from the WMN ("WMN-ICs"; Fig. 2*B*). Each WMN-IC was functionally annotated using multiple linear regression models including both D-prime 2-back and D-prime 0-back performances as well as further covariates as independent variables (Table 1; Ex-

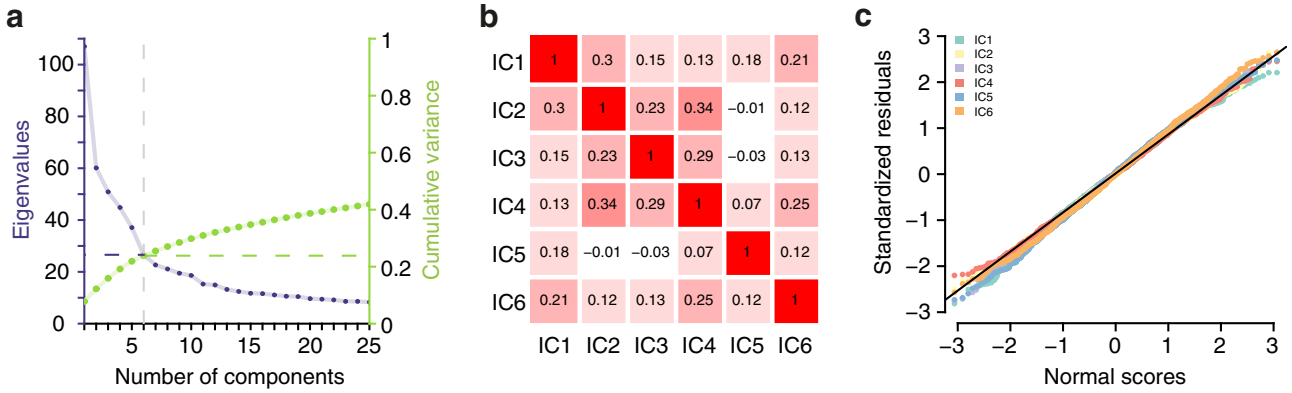

Figure 2. WMN ICA decomposition auxiliary information. ***A***, The eigenvalues (purple, left *y*-axis) and cumulative variance (green, right *y*-axis) of a PCA on the WMN. ***B***, Pearson's correlations between WMN-ICs on the subject-level ($N = 1369$). ***C***, Quantile-quantile plots comparing the standardized residuals (*y*-axis) from multiple linear regression models of each WMN-IC against behavioral measurements and covariates (including the remaining WMN-ICs) with a normal distribution (*x*-axis); D-prime 2-back and D-prime 0-back performances were included as separate predictors in these models. Models with the performance difference of D-prime 2-back and D-prime 0-back as a single predictor yielded highly similar residuals (all $r_{Pearson} > 0.98$).

Table 1. Associations of WMN-ICs with performances

	D-prime 2-back – 0-back (df = 1350)		D-prime 2-back (df = 1349)		D-prime 0-back (df = 1349)		Reaction time 2-back – 0-back (df = 1349)		Episodic memory (df = 1349)	
IC#	β	p	β	p	β	p	β	p	β	p
IC1	0.01	0.85	0.02	0.55	0.03	0.35	-0.01	0.85	0.05	0.12
IC2	0.01	0.69	-0.01	0.74	-0.07	0.02*	-0.02	0.66	0.08	0.006*
IC3	0.24	2.3×10^{-19}***	0.24	2.8×10^{-18}***	-0.15	7.6×10^{-7}***	0.09	0.003*	0.02	0.66
IC4	-0.13	2.8×10^{-6}***	-0.06	0.09	0.25	1.8×10^{-19}***	-0.02	0.52	-0.04	0.18
IC5	-0.01	0.75	-0.01	0.73	0.01	0.85	-0.06	0.09	-0.05	0.19
IC6	-0.05	0.14	-0.04	0.27	0.06	0.12	0.01	0.81	-0.10	9.0×10^{-4}**

The reported p values are FDR corrected (see Materials and Methods); $*p < 0.05$, $**p < 0.001$, $***p < 0.0001$.
The results of the linear models with the WMN-ICs as dependent variables for n-back D-prime performances, n-back reaction time, and episodic memory performance. For the remaining covariates, see Extended data Table 1-1. Voxel-wise associations are described in Extended data Table 1-2. The estimates of statistical power for a voxel-wise analysis and an analysis using WMN-ICs are displayed in Extended data Table 1-3.

tended data Table 1-1; for the distributions of the residuals of the models, see Fig. 2*C*).

Two of the six components were associated with the difference of WM-related and attention-related performances (D-prime 2-back – D-prime 0-back; Table 1). WMN-IC3 was positively associated with the performance difference ($p_{FDR} = 2.3 \times 10^{-19}$, $R^2 = 0.06$) and WMN-IC4 was negatively associated with the performance difference ($p_{FDR} = 2.8 \times 10^{-6}$, $R^2 = 0.02$). We next calculated multiple linear regression models with WM-related performance and attention-related performance as separate predictors. These models were used as main models for all subsequent analyses. In these analyses, WMN-IC3 was significantly associated with both the D-prime 2-back performance ($p_{FDR} = 2.8 \times 10^{-18}$; Fig. 3*B*) and the D-prime 0-back performance with opposite direction of effects ($p_{FDR} = 7.6 \times 10^{-7}$). WMN-IC3 explained 5.8% variance of D-prime 2-back performance, 2.2% variance of D-prime 0-back performance and 0.8% variance of the difference in reaction time between 2-back and 0-back. This component exhibited the most extreme positive loadings ($z > 1.47$, describing the most extreme 10% of absolute values across the WMN-ICs) in bilateral parietal regions, the bilateral middle frontal gyrus, as well as the left precentral gyrus and pars opercularis (Fig. 3*A*; Extended data Fig. 3-1). WMN-IC3 was also associated with sex ($p_{FDR} = 4.0 \times 10^{-6}$); separate analyses for each gender yielded similar results for WM-related and attention-related performances (males: $N = 528$, 2-back performance $R^2 = 0.04$, $p_{FDR} = 4.3 \times 10^{-5}$, 0-back performance $R^2 = 0.03$, $p_{FDR} = 4.0 \times 10^{-4}$, opposite directions of effect; females: $N = 841$, 2-back performance $R^2 = 0.07$, $p_{FDR} = 3.4 \times 10^{-13}$, 0-back performance $R^2 = 0.02$, $p_{FDR} = 0.005$, opposite directions of effect). WMN-IC4 was markedly associated with D-prime 0-back performance ($p_{FDR} = 1.8 \times 10^{-19}$, $R^2 = 0.06$; Fig. 3*D*) but not with D-prime 2-back performance ($p_{FDR} = 0.09$, $R^2 = 0.004$). This component exhibited main posi-

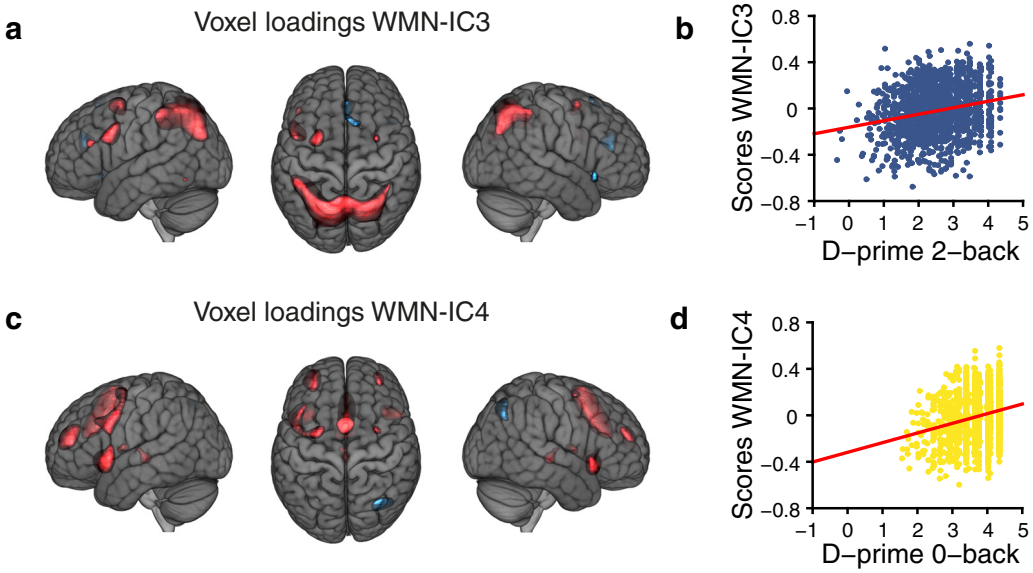

Figure 3. WMN ICA decomposition main findings. Voxel loadings (***A***) of WMN-IC3 and (***C***) of WMN-IC4 illustrated for $|z| > 1.47$ showing the most extreme 10% of the voxel loadings across all WMN-ICs; red depicts positive and blue negative voxel loadings. Associations (***B***) of WMN-IC3 with D-prime 2-back and (***D***) of WMN-IC4 with D-prime 0-back task performances. Annotations of WMN-ICs with anatomic regions are listed in Extended data Figure 3-1. The results of additional WMN ICA decompositions with varying numbers of components are illustrated in Extended data Figures 3-2, 3-3. The brain images are displayed within the MNI152 template and according to neurologic convention.

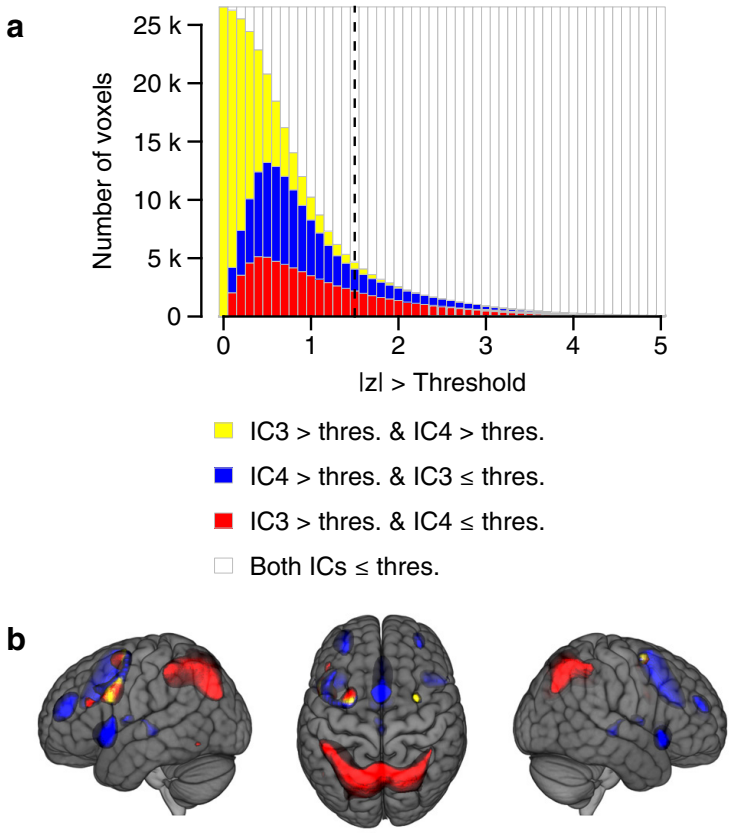

Figure 4. Overlap between WMN-IC3 and WMN-IC4. ***A***, The effects of curtailing the voxel loadings of WMN-IC3 and WMN-IC4 using different thresholds ranging from $|z| > 0$ to $|z| > 5$ (*x*-axis). Stacked bars (*y*-axis) depict the share of all $N = 26542$ voxels that load onto both WMN-IC3 and WMN-IC4 (yellow; i.e., overlap between WMN-IC3 and WMN-IC4), onto WMN-IC4 but not WMN-IC3 (blue), onto WMN-IC3 > thres but not WMN-IC4 ≤ thres (red), and onto neither WMN-IC (white) above the threshold indicated by the *x*-axis. The dashed vertical line highlights the share of voxels loading onto the WMN-ICs above a threshold of $|z| > 1.47$. This threshold includes the most extreme 10% of values across all WMN-ICs and was used for illustrating the brain images and to determine the overlap between WMN-IC3 and WMN-IC4 throughout the paper. ***B***, Brain regions loading with $z > 1.47$ onto both WMN-IC3 and WMN-IC4 (yellow; i.e., overlap between WMN-IC3 and WMN-IC4), only onto WMN-IC4 (blue), and only onto WMN-IC3 (red). The anatomic annotations of clusters loading onto both WMN-IC3 and WMN-IC4 when considering the most extreme 10% of loadings are described in Extended data Figure 4-1. The brain images are displayed within the MNI152 template and according to neurologic convention.

tive loadings bilaterally in frontal regions such as the caudal anterior cingulate gyrus, the insula, and the middle frontal gyrus (Fig. 3C; Extended data Fig. 3-1; focusing on the most extreme 10% of loadings).

The voxel loadings of WMN-IC3 and WMN-IC4 showed only minor overlaps when focusing on the most extreme 10% of loadings, with 1% of all WMN-voxels showing $z > 1.47$ in both WMN-IC3 and WMN-IC4. The overlaps of WMN-IC3 and WMN-IC4 for this threshold, as well as for a range of other thresholds, are illustrated in Figure 4 (yellow color). The overlaps between WMN-IC3 and WMN-IC4 for the most extreme 10% of loadings comprised three distinct clusters of adjacent voxels (Extended data Fig. 4-1). Two of these clusters were located bilaterally in the middle frontal gyrus and the posterior part of the superior frontal gyrus. The third cluster was located in the left superior parietal cortex.

The voxel loadings of the remaining WMN-ICs are shown in Figure 5. Two components showed predominantly lateralized loadings (WMN-IC5 left, WMN-IC6 right) in frontal regions, inferior parietal regions and the cerebellum when focusing on the most extreme 10% of loadings. IC6 was associated with episodic memory performance ($p_{FDR} = 0.0009$, $R^2 = 0.010$); IC5 did not show any FDR-corrected significant associations with task performances. WMN-IC2 loaded bilaterally onto occipital regions like the fusiform gyrus and the lingual gyrus, as well as the cerebellum and the thalamus when considering the most extreme 10% of loadings, and was associated with episodic memory performance ($p_{FDR} = 0.006$, $R^2 = 0.006$) and D-prime 0-back performance ($p_{FDR} = 0.02$, $R^2 = 0.005$). WMN-IC1 loaded bilaterally onto the precuneus, frontal and inferior parietal regions when focusing on the most extreme 10% of loadings and did not show any FDR-corrected significant associations with task performances.

In summary, within the WMN, two out of six networks functionally differentiated between WM performance and attention. A parietally-centered network was mainly associated with WM-related performance and a frontally-centered network was mainly associated with attention-related performance. We verified these results by applying voxel-wise association analyses between the 2-back –

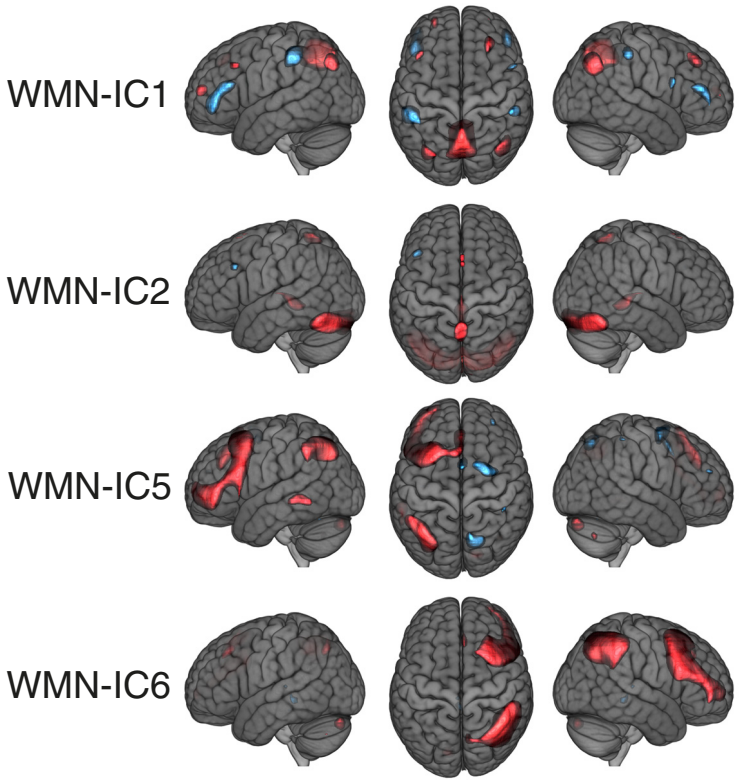

Figure 5. WMN ICA decomposition voxel loadings of the remaining WMN-ICs. The threshold of $|z| > 1.47$ used for illustration displays the most extreme 10% of the voxel loadings across all WMN-ICs; red depicts positive and blue negative voxel loadings. The brain images are displayed within the MNI152 template and according to neurologic convention.

0-back contrast parameter estimates and D-prime 0-back as well as D-prime 2-back performances (Fig. 6*A*,*B*; see also "β performance 0-back" and "β performance 2-back" in NeuroVault; see Extended data Table 1-2 for the remaining variables). On this voxel-wise level, mainly parietal and superior frontal voxels showed positive associations with D-prime 2-back performance and mainly frontal regions showed positive associations with D-prime 0-back performance.

To confirm the stability of the main results from the ICA we applied bootstrapping and cross-validation procedures. The bootstrapping revealed stable network decomposition and robust associations of these networks with task performances in subsamples of $N = 100$ (Fig.

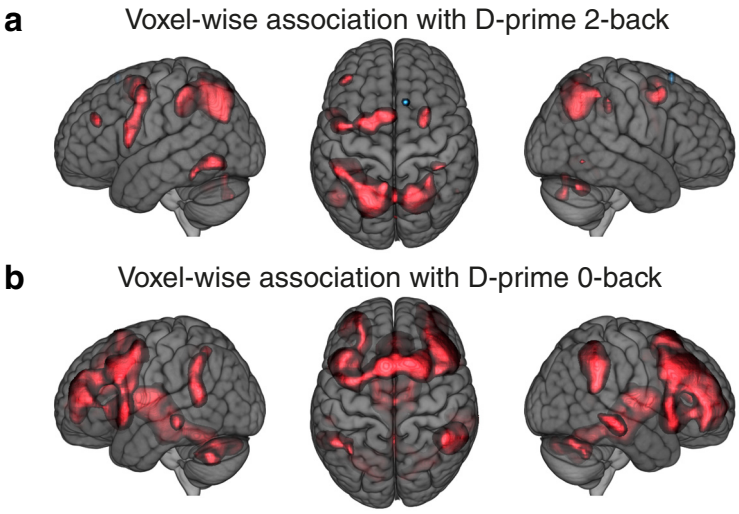

Figure 6. WMN voxel-wise association results. Univariate results for WMN voxels against D-prime 2-back (***A***) and D-prime 0-back (***B***) task performances ($N = 1369$, df $= 1354$); red clusters show FDR-corrected significant positive associations. The brain images are displayed within the MNI152 template and according to neurologic convention.

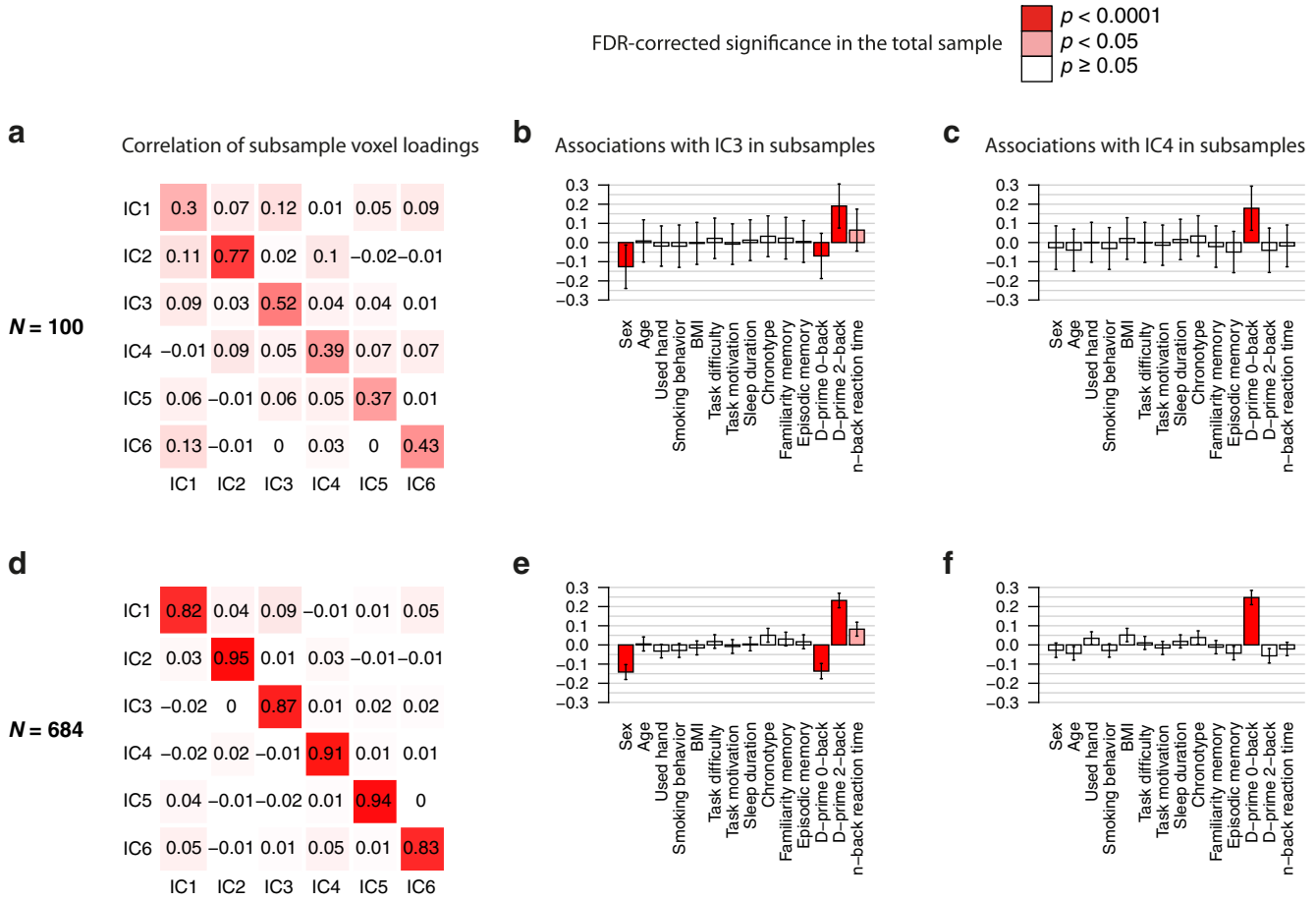

Figure 7. WMN ICA decomposition bootstrapping results. Pearson's correlation coefficients comparing the voxel loadings of ICA decompositions between two nonintersecting subsamples of sizes (***A***) $N = 100$ each and (***D***) $N = 684$ each (i.e., split-halves). Depicted are the averaged correlation coefficients across 100 runs. The associations of WMN-IC3 with task performances and covariates averaged across the 2×100 random subsamples are shown for (***B***) $N = 100$ (df = 85) and (***E***) $N = 684$ (df = 669). The associations of WMN-IC4 with task performances and covariates averaged across the 2×100 random subsamples are shown for (***C***) $N = 100$ (df = 85) and (***F***) $N = 684$ (df = 669). Bars represent the averaged regression coefficients; error bars denote the averaged standard errors of the regression coefficients; red colors in the bar plots describe the FDR-corrected significance of the corresponding WMN-IC's association with the independent variables in the total sample (see top-right legend).

7*A–C*) and of $N = 684$ (i.e., split-half; Fig. 7*D–F*). Cross-validations additionally demonstrated that ICA solutions estimated in a larger subsample could predict task performance in another nonintersecting smaller subsample (WMN-IC3 and D-prime 2-back: averaged $\beta = 0.25$, $p_{nominal} < 0.05$ in 64% of runs; WMN-IC4 and D-prime 0-back: averaged $\beta = 0.23$ $p_{nominal} < 0.05$ in 61% of runs; expected under H_0 is 5%, $p_{empirical} < 0.001$ in both analyses). Additionally, we repeated the ICA decomposition and the association analyses with a varying number of extracted components (between 2 and 10). The results remained very similar when using more than three components (Extended data Figs. 3-2, 3-3). The estimated WMN-ICs from the six-components solution are provided in NeuroVault ("*z* value voxel loadings WMN-IC").

Association of WM-task networks with cortical white matter microstructure

Differences in cortical white matter microstructure impact the activity in functional brain networks (Andrews-Hanna et al., 2007; Burzynska et al., 2011; Palacios et al., 2012; Marstaller et al., 2015). We tested for a global association between white matter microstructure and differences in WMN-IC scores in our sample, separately for each WMN-IC. Cortical white matter microstructure was measured by DTI. We used FA values that are related to fiber orientation (Jones et al., 2013). Data were available for 70 white matter-segmented brain regions in a subsample of 614 subjects from our study. Out of the six networks, the parietally-centered network WMN-IC3 showed a significant global association between white matter microstructure and strength of network activation (Kolmogorov–Smirnov: $D = 0.37$, $p_{FDR} = 1.6 \times 10^{-8}$, for all remaining WMN-ICs $p_{FDR} > 0.24$; Empiric: $p_{FDR} = 0.01$, for all remaining WMN-ICs $p_{FDR} > 0.32$; Fig. 8). The largest positive associations between WMN-IC3 and FA values were found in white matter regions adjacent to the posterior cingulum, the superior parietal cortex, and the precentral gyrus (Table 2).

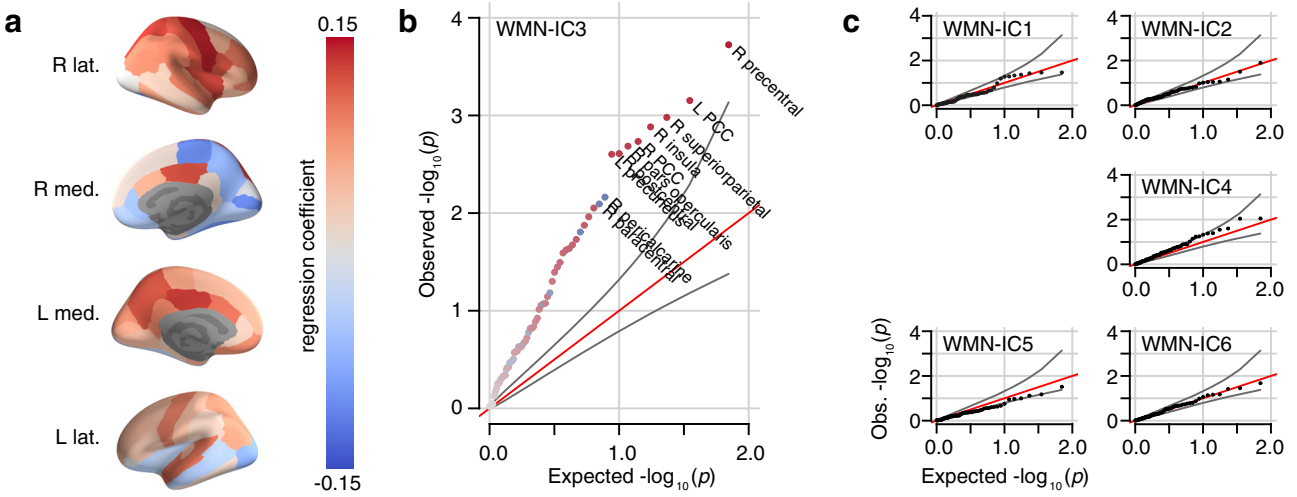

Figure 8. WMN ICA associations with cortical white matter microstructure. ***A***, Associations of WMN-IC3 with the averaged FA values in 70 cortical white matter areas ($N = 614$, df = 602). ***B***, ***C***, Quantile-quantile plots of the -log10(p) values from the linear regressions of averaged FA values against (***B***) WMN-IC3 and (***C***) the remaining five WMN-ICs. The Quantile-quantile plot compares the distribution of -log10(p) values expected at random (x-axis) with the distribution of the observed -log10(p) values (y-axis). Gray curves indicate 95% confidence intervals. Detailed results for all 70 areas are listed in Table 2. R: right; L: left; lat.: lateral; med.: medial.

Comparison of the WM-task networks with external datasets

Functional brain networks can be specifically activated in one given task or can be involved in a variety of different tasks (Cole et al., 2014). To assess the specificity of the WMN-ICs we compared them with results from other studies that cover a wider range of different tasks. We investigated whether the networks derived from the verbal n-back task had previously been identified in others studies using not only the n-back task but also different WM-related paradigms.

We first compared our results with the results of an extensive meta-analysis of neuroimaging studies that includes a number of different WM tasks (Rottschy et al., 2012). The authors reported a "WM core network" of 10 regions that were consistently activated across distinct WM tasks, designs and contrasts. Seven out of these 10 regions overlapped with voxels showing high loadings ($z > 1.47$) in WMN-IC3 or WMN-IC4 derived from our data (Fig. 9): Three of these 10 regions showed high loadings on WMN-IC3 only and three regions showed high loadings on IC4 when focusing on the most extreme 10% of loadings. One region shared high loadings on both WMN-IC3 and WMN-IC4.

Next, we assessed whether the networks identified with our data show similarities with networks derived from NeuroSynth (Yarkoni et al., 2011). NeuroSynth is a meta-analytical brain imaging resource that provides information from 11406 fMRI studies covering a wide range of distinct tasks. Based on a PCA (Fig. 10*A*) and ICA decomposition, we retrieved 16 global networks of brain activations that were found across the included studies and terms (all estimated networks are described in Table 3, Extended data Fig. 10-1; the estimated NeuroSynth-ICs are additionally provided in NeuroVault "z value voxel loadings NeuroSynth-IC"). Two of these global networks (NeuroSynth IC-topic 11 "DLPFC" and IC-topic 8 "parietal") were to a large extent (> 80% of the voxels with $z > 0.70$; $|z| > 0.7$ described the most extreme 10% of absolute values across the NeuroSynth-ICs) located within the WMN derived from our data (Table 3). We compared the loadings of these two networks with the loadings of WMN-ICs of our data. We then retrieved scores of the two NeuroSynth networks for our subjects and associated them with the subjects' task performances. The parietal network (Fig. 10*B*) showed a profound similarity with WMN-IC3 (42% shared variance when comparing voxel loadings within the WMN). The subject-wise scores derived for the NeuroSynth IC-topic parietal were very similar to the scores of our WMN-IC3 ($r_{Pearson} = 0.77$; 59% shared variance; Fig. 10*D*). Correspondingly, WM performance also showed a highly significant association with scores derived for the NeuroSynth IC-topic parietal in our sample (D-prime 2-back: $p_{FDR} = 2.4 \times 10^{-10}$, $R^2 = 0.04$; D-prime 0-back: $p_{FDR} = 0.20$, $R^2 = 0.002$). We did not find a profound similarity of the DLPFC network's voxel loadings (Fig. 10*C*) with any of our WMN-ICs (shared variances < 3.3%). However, the subject-wise scores of the DLPFC network were moderately correlated with the scores of WMN-IC4 ($r_{Pearson} = 0.25$; 6% shared variance; Fig. 10*E*) and were also associated with D-prime 0-back performance in our sample (D-prime 0-back: $p_{FDR} = 1.6 \times 10^{-8}$, $R^2 = 0.04$; D-prime 2-back: $p_{FDR} = 0.45$, $R^2 < 0.001$).

Discussion

Studies on WM related brain activation typically describe a fronto-parietal network being implicated in WM tasks (Klingberg et al., 2002; Owen et al., 2005; Rottschy et al., 2012; Satterthwaite et al., 2013; Constantinidis and Klingberg, 2016; Huang et al., 2016). Based on ICA decomposition we have identified two networks within the WMN that showed distinct functional characteristics. A network with prominent parietal and smaller frontal features was mainly associated with WM-related perfor-

Table 2. Associations of WMN-IC 3 with DTI measurements

White matter-segmented region	Association with WMN-IC3					
	Both hemispheres		Left hemisphere		Right hemisphere	
	β	p	β	p	β	p
Posterior cingulum FA	0.14	0.02*	0.13	0.02*	0.12	0.02*
Precentral FA	0.13	0.02*	0.09	0.10	0.14	0.02*
Superiorparietal FA	0.12	0.02*	0.07	0.20	0.13	0.02*
Superiortemporal FA	0.11	0.05	0.10	0.07	0.09	0.09
Pars opercularis FA	0.09	0.08	0.03	0.54	0.12	0.02*
Postcentral FA	0.09	0.08	0.04	0.44	0.12	0.02*
Caudal anterior cingulum FA	0.09	0.08	0.10	0.06	0.06	0.32
Inferiorparietal FA	0.09	0.10	0.07	0.23	0.08	0.12
Rostralmiddlefrontal FA	0.08	0.11	0.05	0.40	0.09	0.08
Transversetemporal FA	0.07	0.17	0.05	0.40	0.07	0.16
Fusiform gyrus FA	−0.07	0.18	−0.03	0.63	−0.09	0.08
Insula FA	0.07	0.19	−0.03	0.61	0.12	0.02
Supramarginal gyrus FA	0.06	0.24	0.02	0.67	0.09	0.10
Isthmus of cingulum FA	0.06	0.25	0.09	0.10	0.02	0.77
Caudalmiddlefrontal FA	0.06	0.32	0.04	0.44	0.06	0.32
Pars triangularis FA	0.05	0.36	0.03	0.54	0.05	0.38
Inferiortemporal FA	0.05	0.38	0.02	0.73	0.06	0.24
Pars orbitalis FA	0.05	0.38	0.00	0.96	0.09	0.10
Cuneus FA	0.05	0.38	0.08	0.14	−0.01	0.96
Entorhinal FA	0.04	0.44	0.06	0.32	0.01	0.83
Lateral occipital FA	−0.04	0.49	−0.07	0.19	0.00	0.96
Medial orbitofrontal FA	0.04	0.50	0.08	0.13	−0.04	0.50
Superiorfrontal FA	0.03	0.53	0.05	0.38	0.01	0.83
Paracentral FA	−0.03	0.55	0.07	0.23	−0.10	0.06
Precuneus FA	0.03	0.56	0.11	0.02	−0.07	0.23
Temporal pole FA	−0.03	0.62	0.00	0.96	−0.05	0.39
Pericalcarine FA	−0.03	0.64	0.06	0.30	−0.10	0.06
Rostral anterior cingulum FA	−0.02	0.68	0.02	0.64	−0.05	0.35
Frontal pole FA	0.02	0.73	−0.02	0.67	0.06	0.28
Parahippocampal FA	−0.01	0.81	0.03	0.61	−0.05	0.40
Corpus callosum FA	−0.01	0.84	0.00	0.96	−0.02	0.67
Middletemporal FA	−0.01	0.96	−0.04	0.50	0.03	0.62
Banks of superior temporal sulcus FA	0.00	0.96	−0.03	0.55	0.03	0.61
Lateral orbitofrontal FA	0.00	0.96	−0.04	0.51	0.05	0.40
Lingual gyrus FA	0.00	0.98	0.04	0.44	−0.04	0.49

The reported p values are FDR corrected ($\alpha < 0.05$) for three (both hemispheres, left hemisphere, right hemisphere) × 70 (anatomic regions) tests; $*p < 0.05$. All df = 602.
Shown are FDR-corrected p values and regression coefficients describing the associations of FA measures (averaged across both hemispheres, for the left hemisphere, and for the right hemisphere) with the estimates of WMN-IC3.

mance (5.8% variance explained), the associations with attention-related performance were smaller (2.2% variance explained) and in the opposite direction of effect. A second network of predominantly frontal areas (left DLPFC, ACC, both insulae) was merely relevant for attention-related behavior (6.2% variance explained).

Our findings of a frontally-centered and a parietally-centered network involved in different aspects of WM-

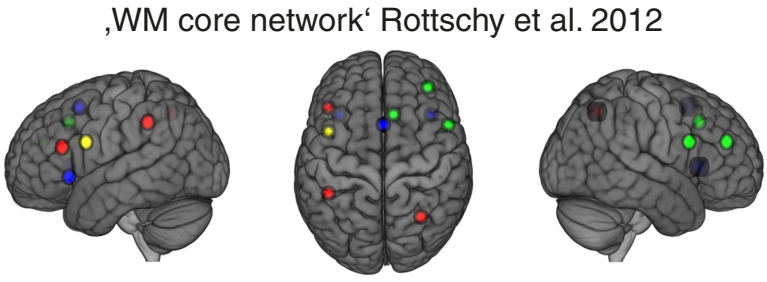

Figure 9. WMN-IC3 and WMN-IC4 in comparison with a WM core network described in Rottschy et al. (2012). Red regions overlap with WMN-IC3 ($z > 1.47$), blue regions overlap with WMN-IC4 ($z > 1.47$), yellow region overlaps with WMN-IC3 and WMN-IC4; the green regions do not overlap with WMN-IC3 or WMN-IC4. The brain images are displayed within the MNI152 template and according to neurologic convention.

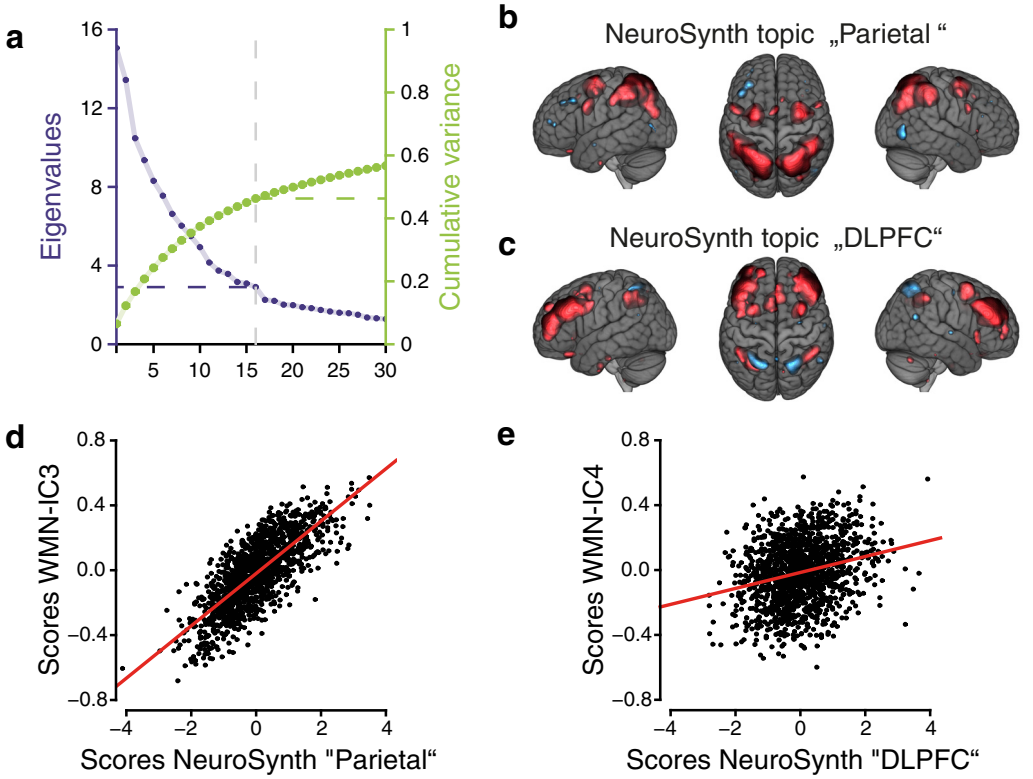

Figure 10. WMN-IC3 and WMN-IC4 compared to an ICA decomposition derived from NeuroSynth. ***A***, The eigenvalues (purple, left *y*-axis) and cumulative variance (green, right *y*-axis) of the PCA on voxel loadings of 233 NeuroSynth terms. Voxel loadings (***B***) of the NeuroSynth IC-topic parietal (IC8) and voxel loadings (***C***) of the NeuroSynth IC-topic DLPFC (IC11) with $|z| > 0.70$. ***D***, ***E***, Comparison of the subject-wise scores of the WMN-IC3 and WMN-IC4 with the subject-wise scores of the NeuroSynth IC-topic eight parietal ($r_{Pearson} = 0.76$) and IC-topic 11 DLPFC ($r_{Pearson} = 0.25$). All 16 estimated NeuroSynth IC-topics are shown in Extended data Figure 10-1. The brain images are displayed within the MNI152 template and according to radiologic convention.

task performances are in line with recent ROI-based studies that have reported distinct functional roles of frontal and parietal regions on WM (Kundu et al., 2015; Sarma et al., 2016; Li et al., 2017). The functional results from our estimated networks were also consistent with the voxel-wise results from our data. Notably, using ICA decomposition to estimate brain networks resulted in several advantages as compared to voxel-wise or ROI-based analyses. Both voxel-wise and ROI-based approaches require prior knowledge for defining brain activation patterns relevant for performance, either regarding the subject's task performance or the anatomic ROIs. In contrast, the ICA decompositions applied here estimated brain activation networks based on the WMN contrast estimates and did neither include performance measures into the estimation nor preselect voxels based on prior assumptions. Thus, the WMN-ICs constitute data-driven and unbiased measures of brain networks that underlie the task performances. Importantly, ICA decomposition optimized the detection rate of true effects for associating brain activation with WM task performance by considerably decreasing the number of tests performed, from $N =$ 26,542 voxel-wise tests to 6 association analyses with the WMN-ICs, effectively reducing the false-positive rate and increasing statistical power. Furthermore, ICA decomposition enabled us to represent brain networks that were statistically maximally independent. By using the subject's performance measurements, we could show that these networks exhibit distinct functional characteristics. Subjects with high scores on a WMN-IC showed increased coactivation of the voxels that loaded highly onto this WMN-IC, we thus interpreted WMN-ICs as networks of coactivating brain regions. The identification of distinct functional networks within the WM brain activation is in line with numerous recent studies demonstrating that the brain activation at rest as well as during different tasks is most likely based on distinct but possibly spatially overlapping networks (Power et al., 2011; Cole et al., 2014, 2016; Xu et al., 2016). In contrast, univariate voxel-wise analyses or ROI-based approaches would not allow to identify data-driven and statistically independent subnetworks of brain activation that underlie the brain activation during the WM task. Importantly, due to the large sample size used here we can provide robust network estimates that can also be applied to samples with smaller sample sizes.

Sets of brain regions that appear similar to our parietally-centered network have been described in past studies as orienting system for visual events (Fan and Posner, 2004) or dorsal attention network (Power et al., 2011; Petersen and Posner, 2012). The frontally-centered network derived from our data resembles the cingulo-opercular network that has been linked to maintaining alertness (Coste and Kleinschmidt, 2016). The two net-

Table 3. Description of ICs estimated from NeuroSynth data

NeuroSynth IC-topic	IC#	The 10 most-contributing NeuroSynth terms	% voxels in WMN
DLPFC	11	Dorsolateral; dorsolateral_prefrontal; dlpfc; cortex_dlpfc; working; working_memory; prefrontal; prefrontal_cortex; executive; load	88%
Parietal	8	Intraparietal; intraparietal_sulcus; parietal_cortex; parietal; posterior_parietal; superior_parietal; spatial; fronto_parietal; attentional; sulcus	83%
Morphometry versus demand	9	Voxel; matter; morphometry; voxel_morphometry; demands; volume; task; difficulty; working; working_memory	71%
Inferior frontal	12	Inferior_frontal; semantic; word; inferior; language; frontal_gyrus; words; sentence; meaning; sentences	59%
Fusiform gyrus	7	Fusiform; fusiform_gyrus; face; objects; faces; recognition; category; object; visual; occipital	54%
Motion/observation	13	Motion; body; observation; viewed; perception; actions; visual; occipital_cortex; direction; viewing	52%
Motor cortex	15	Motor; movement; motor_cortex; primary_motor; hand; finger; movements; premotor; sensorimotor; supplementary_motor	51%
Sensory system	5	Secondary; somatosensory; pain; somatosensory_cortex; stimulation; insular; insula; primary; sensory; intensity	42%
Basal ganglia	2	Basal_ganglia; ganglia; basal; putamen; subcortical; thalamus; striatal; caudate; striatum; nucleus	42%
Temporal	4	Superior_temporal; superior; auditory; speech; temporal_gyrus; temporal_sulcus; temporal; posterior_superior; linguistic; gyrus	40%
ACC	16	Anterior_cingulate; anterior; acc; cingulate_cortex; cingulate; cortex_acc; dorsal_anterior; anterior_insula; insula; cortex_anterior	40%
Striatum	10	Ventral_striatum; reward; striatum; ventral; value; nucleus; striatal; decision_making; orbitofrontal; orbitofrontal_cortex	38%
Medial prefrontal	6	Social; medial_prefrontal; junction; theory; temporo; medial; states; person; mental; prefrontal_cortex	35%
Default mode	3	Default_mode; mode; default; mode_network; resting; resting_state; state; posterior_cingulate; independent_component; functional_connectivity	33%
Hippocampus	14	Hippocampal; medial_temporal; hippocampus; parahippocampal; temporal_lobe; encoding; episodic; episodic_memory; parahippocampal_gyrus; lobe	20%
Amygdala	1	Neutral; amygdala; emotion; fear; emotional; expressions; facial; affective; emotions; anxiety	14%

IC-topics were assigned based on NeuroSynth terms with the highest loading. For voxels showing the highest loadings ($z > 0.70$), we calculated the percentage of voxels being located within the WMN (% voxels in WMN).

works identified in our study spatially overlap in three separate clusters when focusing on the most extreme 10% of loadings. Two of these clusters were located bilaterally in the middle frontal gyrus and the posterior part of the superior frontal gyrus. The third cluster was located in the left superior parietal cortex. Overlaps between brain networks could represent regions of convergence between otherwise segregated functional networks (Sporns, 2013). Links between distinct networks are presumably features of brain organization and important for complex behaviors (Yeo et al., 2014). Accordingly, the lateral PFC (which includes the middle frontal gyrus) has been proposed to serve as a globally connected functional hub that is involved in cognitive control (Cole et al., 2012). Together, the most extreme 10% of voxel loadings of the two networks relevant for WM task performance in our study closely overlap with a WM core network identified in an extensive meta-analysis of WM neuroimaging studies by Rottschy et al. (2012) that included a number of other WM tasks besides the verbal n-back task used here. Importantly, both of the two networks estimated in our study overlap with distinct parts of this global WM core network. Furthermore, the parietally-centered network identified in our study sample showed considerable similarity with a parietal network derived from NeuroSynth (Yarkoni et al., 2011). This parietal network derived from NeuroSynth was estimated across a large body of results from neuroimaging studies using many different paradigms. These results imply that especially the parietally-centered network, which was associated with WM-related task performance in our sample, is an important and stable network implicated in WM-related cognitive functioning.

This parietally-centered network was furthermore associated with global differences in FA estimates in our subjects. FA describes aspects of white matter microstructure related to fiber orientation (Jones et al., 2013) and can be modulated by myelination (Beaulieu, 2002). Measurements of FA have been observed to decrease with increasing age (Inano et al., 2011) and after moderate to severe traumatic brain injury (Kraus et al., 2007). Properties of white matter microstructure have also been shown to affect large-scale functional networks such as the default mode network (Andrews-Hanna et al., 2007), the WMN (Burzynska et al., 2011; Palacios et al., 2012; Darki and Klingberg, 2015), the salience network, and the fronto-parietal network (Marstaller et al., 2015). The parietally-centered network in our study was globally associated with FA measures across the white matter-segmented regions. Conversely, the other networks esti-

mated here did not show any FA-associations. Positive associations of FA with fMRI measurements or with connectivity measures have been proposed to represent better transmission and stronger functional connections (Warbrick et al., 2017). FA measures in fronto-parietal tracts have moreover been associated with WM performance (Charlton et al., 2010; Vestergaard et al., 2011; Darki and Klingberg, 2015). A recent large-scale study of $N = 1584$ subjects reported that functional connectivity between brain regions was influenced by lesions in white matter tracts directly connecting the brain regions, but also by white matter load in other, not directly connected tracts (Langen et al., 2017). Thus, global white matter integrity might contribute to the WM performance-relevant coactivation observed in our study. Additionally, we observed that FA measures of single white matter-segmented regions adjacent to the parietally-centered network's cortical main foci (specifically the posterior cingulum, superior parietal, and precentral regions) were associated with coactivation within the network.

WM and attention are closely related neurocognitive domains (Eriksson et al., 2015; Constantinidis and Klingberg, 2016). Importantly, these neurocognitive domains are also affected in neuropsychiatric disorders like schizophrenia (Barch and Ceaser, 2012). A meta-analysis across 41 neuroimaging studies observed reduced activation of the left DLPFC and the ACC in schizophrenia patients during executive tasks (Minzenberg et al., 2009). Barch and Ceaser (2012) depicted that the robustly observed altered DLPFC activation in schizophrenia could either directly impact cognitive functions such as WM or interfere with top-down functions such as proactive control that in turn mediate the effect on WM. Our observation that a network of frontal regions including the DLPFC and ACC was mainly associated with attention-related performance coincides with the assumption of impaired general executive functions rather than isolated WM function in schizophrenia. Other studies investigating cognitive deficits in schizophrenia have come to similar conclusions of a deficit in general cognitive ability in schizophrenia (Haut et al., 2015).

To summarize, we have identified two networks within the WMN that showed distinct functional characteristics with respect to attention-related and WM-related task performances. Compared to voxel-wise analyses, using a multivariate approach led to more specific results with higher effect sizes and higher statistical power while minimizing the burden of multiple testing. Low statistical power in combination with a large number of statistical tests is a prevalent source of critique regarding the existing neuroimaging literature (Poldrack et al., 2017; Szucs and Ioannidis, 2017), especially in combination with multiple high-dimensional datasets such as imaging genetic studies (Bigos and Weinberger, 2010; Medland et al., 2014; Poline et al., 2015). Van Snellenberg et al. (2016) have stressed that finding replicable biomarkers of WM will help to broaden our understanding of the associated neural, molecular or genetic mechanisms. Our findings take a step in this direction by providing stable network estimates for application in independent samples (http://neurovault.org/collections/EYCSLZUZ/). This allows future studies to investigate functional distinct brain networks that are implicated in human cognition.

References

Alderson RM, Kasper LJ, Hudec KL, Patros CHG (2013) Attention-deficit/hyperactivity disorder (ADHD) and working memory in adults: a meta-analytic review. Neuropsychology 27:287–302. CrossRef

Andrews-Hanna JR, Snyder AZ, Vincent JL, Lustig C, Head D, Raichle ME, Buckner RL (2007) Disruption of large-scale brain systems in advanced aging. Neuron 56:924–935. CrossRef Medline

Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38:95–113. CrossRef Medline

Baddeley A (2012) Working memory: theories, models, and controversies. Annu Rev Psychol 63:1–29. CrossRef Medline

Barch DM, Ceaser A (2012) Cognition in schizophrenia: core psychological and neural mechanisms. Trends Cogn Sci 16:27–34. CrossRef Medline

Basser PJ (1995) Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. NMR Biomed 8:333–344. Medline

Beaulieu C (2002) The basis of anisotropic water diffusion in the nervous system - a technical review. NMR Biomed 15:435–455. CrossRef Medline

Bigos KL, Weinberger DR (2010) Imaging genetics - days of future past. Neuroimage 53:804–809. CrossRef Medline

Burzynska AZ, Nagel IE, Preuschhof C, Li S-C, Lindenberger U, Bäckman L, Heekeren HR (2011) Microstructure of frontoparietal connections predicts cortical responsivity and working memory performance. Cereb Cortex 21:2261–2271. CrossRef Medline

Charlton RA, Barrick TR, Lawes INC, Markus HS, Morris RG (2010) White matter pathways associated with working memory in normal aging. Cortex 46:474–489. CrossRef

Chiappetta P, Roubaud MC, Torrésani B (2004) Blind source separation and the analysis of microarray data. J Comput Biol 11:1090–1109. CrossRef Medline

Cole MW, Yarkoni T, Repovs G, Anticevic A, Braver TS (2012) Global connectivity of prefrontal cortex predicts cognitive control and intelligence. J Neurosci 32:8988–8999. CrossRef Medline

Cole MW, Bassett DS, Power JD, Braver TS, Petersen SE (2014) Intrinsic and task-evoked network architectures of the human brain. Neuron 83:238–251. CrossRef Medline

Cole MW, Ito T, Bassett DS, Schultz DH (2016) Activity flow over resting-state networks shapes cognitive task activations. Nat Neurosci 19:1718–1726. CrossRef

Constantinidis C, Klingberg T (2016) The neuroscience of working memory capacity and training. Nat Rev Neurosci 17:438–449. CrossRef Medline

Coste CP, Kleinschmidt A (2016) Cingulo-opercular network activity maintains alertness. Neuroimage 128:264–272. CrossRef Medline

Darki F, Klingberg T (2015) The role of fronto-parietal and fronto-striatal networks in the development of working memory: a longitudinal study. Cereb Cortex 25:1587–1595. CrossRef Medline

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31:968–980. CrossRef Medline

Engreitz JM, Daigle BJ, Marshall JJ, Altman RB (2010) Independent component analysis: mining microarray data for fundamental human gene expression modules. J Biomed Inform 43:932–944. CrossRef Medline

Eriksson J, Vogel EK, Lansner A, Bergström F, Nyberg L (2015) Neurocognitive architecture of working memory. Neuron 88:33–46. CrossRef Medline

Fan J, Posner M (2004) Human attentional networks. Psychiatr Prax 31:210–214. CrossRef

Fischl B (2012) FreeSurfer. Neuroimage 62:774–781. CrossRef Medline

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, Van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002) Whole brain segmentation: neurotechnique automated labeling of neuroanatomical structures in the human brain. Neuron 33:341–355. CrossRef

Forbes NF, Carrick LA, McIntosh AM, Lawrie SM (2009) Working memory in schizophrenia: a meta-analysis. Psychol Med 39:889–905. CrossRef Medline

Gorgolewski KJ, Varoquaux G, Rivera G, Schwartz Y, Ghosh SS, Maumet C, Sochat VV, Nichols TE, Poldrack RA, Poline J-B, Yarkoni T, Margulies DS (2016) NeuroVault.org: a repository for sharing unthresholded statistical maps, parcellations, and atlases of the human brain. Neuroimage 124:1242–1244. CrossRef Medline

Haut KM, Karlsgodt KH, Bilder RM, Congdon E, Freimer NB, London ED, Sabb FW, Ventura J, Cannon TD (2015) Memory systems in schizophrenia: modularity is preserved but deficits are generalized. Schizophr Res 168:223–230. CrossRef Medline

Heck A, Fastenrath M, Ackermann S, Auschra B, Bickel H, Coynel D, Gschwind L, Jessen F, Kaduszkiewicz H, Maier W, Milnik A, Pentzek M, Riedel-Heller SG, Ripke S, Spalek K, Sullivan P, Vogler C, Wagner M, Weyerer S, Wolfsgruber S, de Quervain DJ-F, Papassotiropoulos A (2014) Converging genetic and functional brain imaging evidence links neuronal excitability to working memory, psychiatric disease, and brain activity. Neuron 81:1203–1213. CrossRef Medline

Huang AS, Klein DN, Leung H-C (2016) Load-related brain activation predicts spatial working memory performance in youth aged 9–12 and is associated with executive function at earlier ages. Dev Cogn Neurosci 17:1–9. CrossRef Medline

Hyvärinen A, Oja E (2000) Independent component analysis: algorithms and applications. Neural Netw 13:411–430. Medline

Inano S, Takao H, Hayashi N, Abe O, Ohtomo K (2011) Effects of age and gender on white matter integrity. Am J Neuroradiol 32:2103–2109. CrossRef Medline

Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) FSL. Neuroimage 62:782–790. CrossRef

Jones DK, Knösche TR, Turner R (2013) White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. Neuroimage 73:239–254. CrossRef Medline

Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang M-C, Christensen GE, Collins DL, Gee J, Hellier P, Song JH, Jenkinson M, Lepage C, Rueckert D, Thompson P, Vercauteren T, Woods RP, Mann JJ, Parsey RV (2009) Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 46:786–802. CrossRef Medline

Klingberg T, Forssberg H, Westerberg H (2002) Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. J Cogn Neurosci 14:1–10. CrossRef Medline

Kong W, Vanderburg CR, Gunshin H, Rogers JT, Huang X (2008) A review of independent component analysis application to microarray gene expression data. Biotechniques 45:501–520. CrossRef Medline

Kraus MF, Susmaras T, Caughlin BP, Walker CJ, Sweeney JA, Little DM (2007) White matter integrity and cognition in chronic traumatic brain injury: a diffusion tensor imaging study. Brain 130:2508–2519. CrossRef Medline

Kundu B, Chang J-Y, Postle BR, Van Veen BD (2015) Context-specific differences in fronto-parieto-occipital effective connectivity during short-term memory maintenance. Neuroimage 114:320–327. CrossRef Medline

Lang PJ, Bradley MM, Cuthbert BN (2008) International affective picture system (IAPS): affective ratings of pictures and instruction manual. Technical report A-8. Gainesville, FL: University of Florida. Available at http://www.unifesp.br/dpsicobio/adap/instructions.pdf.

Langen CD, Zonneveld HI, White T, Huizinga W, Cremers LGM, de Groot M, Ikram MA, Niessen WJ, Vernooij MW (2017) White matter lesions relate to tract-specific reductions in functional connectivity. Neurobiol Aging 51:97–103. CrossRef Medline

Lee J, Park S (2005) Working memory impairments in schizophrenia: a meta-analysis. J Abnorm Psychol 114:599–611. CrossRef Medline

Li S, Cai Y, Liu J, Li D, Feng Z, Chen C, Xue G (2017) Dissociated roles of the parietal and frontal cortices in the scope and control of attention during visual working memory. Neuroimage 149:210–219. CrossRef Medline

Macmillan NA, Creelman CD (1990) Response bias: characteristics of detection theory, threshold theory, and "nonparametric" indexes. Psychol Bull 107:401–413. CrossRef

Marazziti D, Consoli G, Picchetti M, Carlini M, Faravelli L (2010) Cognitive impairment in major depression. Eur J Pharmacol 626:83–86. CrossRef Medline

Marstaller L, Williams M, Rich A, Savage G, Burianová H (2015) Aging and large-scale functional networks: white matter integrity, gray matter volume, and functional connectivity in the resting state. Neuroscience 290:369–378. CrossRef Medline

Medland SE, Jahanshad N, Neale BM, Thompson PM (2014) Whole-genome analyses of whole-brain data: working within an expanded search space. Nat Neurosci 17:791–800. CrossRef Medline

Mesholam-Gately RI, Giuliano AJ, Goff KP, Faraone SV, Seidman LJ (2009) Neurocognition in first-episode schizophrenia: a meta-analytic review. Neuropsychology 23:315–336. CrossRef Medline

Miller KM, Price CC, Okun MS, Montijo H, Bowers D (2009) Is the N-back task a valid neuropsychological measure for assessing working memory? Arch Clin Neuropsychol 24:711–717. CrossRef Medline

Minzenberg MJ, Laird AR, Thelen S, Carter CS, Glahn DC (2009) Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia. Arch Gen Psychiatry 66:811–822. CrossRef

Modat M, Ridgway GR, Taylor ZA, Lehmann M, Barnes J, Hawkes DJ, Fox NC, Ourselin S (2010) Fast free-form deformation using graphics processing units. Comput Methods Programs Biomed 98:278–284. CrossRef Medline

Murdoch DJ, Tsai Y-L, Adcock J (2008) P-values are random variables. Am Stat 62:242–245. CrossRef

Muschelli J, Sweeney E, Lindquist M, Crainiceanu C (2015) fslr: connecting the FSL software with R. R J 7:163–175. Medline

Nagel BJ, Barlett VC, Schweinsburg AD, Tapert SF (2005) Neuropsychological predictors of BOLD response during a spatial working memory task in adolescents: what can performance tell us about fMRI response patterns? J Clin Exp Neuropsychol 27:823–839. CrossRef Medline

Owen AM, McMillan KM, Laird AR, Bullmore E (2005) N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum Brain Mapp 25:46–59. CrossRef Medline

Palacios EM, Sala-Llonch R, Junque C, Roig T, Tormos JM, Bargallo N, Vendrell P (2012) White matter integrity related to functional working memory networks in traumatic brain injury. Neurology 78:852–860. CrossRef Medline

Petersen SE, Posner MI (2012) The attention system of the human brain: 20 years after. Annu Rev Neurosci 35:73–89. CrossRef Medline

Poldrack RA, Baker CI, Durnez J, Gorgolewski KJ, Matthews PM, Munafò MR, Nichols TE, Poline J-B, Vul E, Yarkoni T (2017) Scanning the horizon: towards transparent and reproducible neuroimaging research. Nat Rev Neurosci 18:115–126. CrossRef Medline

Poline J-B, Breeze JL, Frouin V (2015) Imaging Genetics with fMRI. In: fMRI: from nuclear spins to brain functions (Uludag L, Ugurbil K, Berliner L, eds), pp. 699–738 New York, NY: Springer.

Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE

(2011) Functional network organization of the human brain. Neuron 72:665–678. CrossRef Medline

Rottschy C, Langner R, Dogan I, Reetz K, Laird AR, Schulz JB, Fox PT, Eickhoff SB (2012) Modelling neural correlates of working memory: a coordinate-based meta-analysis. Neuroimage 60:830–846. CrossRef Medline

Sarma A, Masse NY, Wang X-J, Freedman DJ (2016) Task-specific versus generalized mnemonic representations in parietal and prefrontal cortices. Nat Neurosci 19:143–149. CrossRef

Satterthwaite TD, Wolf DH, Erus G, Ruparel K, Elliott MA, Gennatas ED, Hopson R, Jackson C, Prabhakaran K, Bilker WB, Calkins ME, Loughead J, Smith A, Roalf DR, Hakonarson H, Verma R, Davatzikos C, Gur RC, Gur RE (2013) Functional maturation of the executive system during adolescence. J Neurosci 33:16249–16261. CrossRef Medline

Seeley WW, Crawford RK, Zhou J, Miller BL, Greicius MD (2009) Neurodegenerative diseases target large-scale human brain networks. Neuron 62:42–52. CrossRef Medline

Sporns O (2013) Network attributes for segregation and integration in the human brain. Curr Opin Neurobiol 23:162–171. CrossRef Medline

Szucs D, Ioannidis JPA (2017) Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. PLoS Biol 15:e2000797. CrossRef

Todd JJ, Marois R (2004) Capacity limit of visual short-term memory in human posterior parietal cortex. Nature 428:751–754. CrossRef Medline

Todd JJ, Marois R (2005) Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. Cogn Affect Behav Neurosci 5:144–155. Medline

Ullman H, Almeida R, Klingberg T (2014) Structural maturation and brain activity predict future working memory capacity during childhood development. J Neurosci 34:1592–1598. CrossRef Medline

Van Snellenberg JX, Girgis RR, Horga G, van de Giessen E, Slifstein M, Ojeil N, Weinstein JJ, Moore H, Lieberman JA, Shohamy D, Smith EE, Abi-Dargham A (2016) Mechanisms of working memory impairment in schizophrenia. Biol Psychiatry 80:617–626. CrossRef Medline

Vestergaard M, Madsen KS, Baaré WFC, Skimminge A, Ejersbo LR, Ramsøy TZ, Gerlach C, Åkeson P, Paulson OB, Jernigan TL (2011) White matter microstructure in superior longitudinal fasciculus associated with spatial working memory performance in children. J Cogn Neurosci 23:2135–2146. CrossRef

Wager TD, Smith EE (2003) Neuroimaging studies of working memory: a meta-analysis. Cogn Affect Behav Neurosci 3:255–274. Medline

Warbrick T, Rosenberg J, Shah NJ (2017) The relationship between BOLD fMRI response and the underlying white matter as measured by fractional anisotropy (FA): a systematic review. Neuroimage 153:369–381. CrossRef Medline

Whitcher B, Schmid VJ, Thornton A (2011) Working with the DICOM and NIfTI data standards in R. J Stat Softw 44:1–28. CrossRef

Xu J, Potenza MN, Calhoun VD, Zhang R, Yip SW, Wall JT, Pearlson GD, Worhunsky PD, Garrison KA, Moran JM (2016) Large-scale functional network overlap is a general property of brain functional organization: reconciling inconsistent fMRI findings from general-linear-model-based analyses. Neurosci Biobehav Rev 71:83–100. CrossRef Medline

Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD (2011) Large-scale automated synthesis of human functional neuroimaging data. Nat Methods 8:665–670. CrossRef Medline

Yeo BTT, Krienen FM, Chee MWL, Buckner RL (2014) Estimates of segregation and overlap of functional connectivity networks in the human cerebral cortex. Neuroimage 88:212–227. CrossRef

Zhou J, Gennatas ED, Kramer JH, Miller BL, Seeley WW (2012) Predicting regional neurodegeneration from the healthy brain functional connectome. Neuron 73:1216–1227. CrossRef Medline

Zou Q, Ross TJ, Gu H, Geng X, Zuo X-N, Hong LE, Gao J-H, Stein EA, Zang Y-F, Yang Y (2013) Intrinsic resting-state activity predicts working memory brain activation and behavioral performance. Hum Brain Mapp 34:3204–3215. CrossRef Medline

4.2 Exhaustive search for epistatic effects on the human methylome

Egli, T., Vukojevic, V., Sengstag, T., Jacquot, M., Cabezón, R., Coynel, D., Freytag, V., Heck, A., Vogler, C., de Quervain, D. J.-F., Papassotiropoulos, A., & Milnik, A. (2017). *Scientific Reports, 7*, 13669

SCIENTIFIC REPORTS

OPEN

Exhaustive search for epistatic effects on the human methylome

Tobias Egli[1,2], Vanja Vukojevic[1,2,4], Thierry Sengstag[6,7], Martin Jacquot[6], Rubén Cabezón[6], David Coynel[2,5], Virginie Freytag[1,2], Angela Heck[1,2,3], Christian Vogler[1,2,3], Dominique J.-F. de Quervain[2,3,5], Andreas Papassotiropoulos[1,2,3,4] & Annette Milnik[1,2,3]

Received: 9 May 2017
Accepted: 22 September 2017
Published online: 20 October 2017

Studies assessing the existence and magnitude of epistatic effects on complex human traits provide inconclusive results. The study of such effects is complicated by considerable increase in computational burden, model complexity, and model uncertainty, which in concert decrease model stability. An additional source introducing significant uncertainty with regard to the detection of robust epistasis is the biological distance between the genetic variation and the trait under study. Here we studied CpG methylation, a genetically complex molecular trait that is particularly close to genomic variation, and performed an exhaustive search for two-locus epistatic effects on the CpG-methylation signal in two cohorts of healthy young subjects. We detected robust epistatic effects for a small number of CpGs ($N = 404$). Our results indicate that epistatic effects explain only a minor part of variation in DNA-CpG methylation. Interestingly, these CpGs were more likely to be associated with gene-expression of nearby genes, as also shown by their overrepresentation in DNase I hypersensitivity sites and underrepresentation in CpG islands. Finally, gene ontology analysis showed a significant enrichment of these CpGs in pathways related to HPV-infection and cancer.

Statistical epistasis describes a higher-order dependency in which the effect of a single-locus genotype depends on the genotype at another locus, a phenomenon also called statistical interaction[1]. There is hitherto little evidence for robust and replicable epistatic effects on complex human traits[1], although epistasis is often used as a potential explanation for missing heritability or for the instability of main effects in genetic association studies[2–4]. It has been suggested that higher-order dependencies will explain only a minor part of complex phenotypic variation, in comparison to independent additive genetic effects[5,6].

Genetic variation (e.g. single nucleotide polymorphisms, SNPs) and DNA-CpG methylation can be investigated at high resolution and throughput, which allows a hypothesis-free and exhaustive screening for epistatic effects on a genome-wide scale. However, the study of such epistatic effects is complicated by considerable increase in computational burden[7,8], model complexity, and model uncertainty, which in concert decrease model stability[9–11]. The success of these association analyses crucially relies on the expected effect size, a suitable sample size and the availability of a well-matched replication study[4,7].

For the majority of complex human traits, such as neuropsychiatric diseases, only small effect sizes can be expected for single genomic loci[12–15]. Hence, such complex human traits and diseases[16] are not well amenable to screening for epistatic effects. Gene expression also represents a genetically complex trait, however it is functionally closely related to the DNA sequence variation[17]. Two-locus epistatic effects have been reported for gene expression[1,18,19], although part of these effects might be due to spurious associations mediated by main effects of SNPs[20].

To further investigate the relevance of two-locus epistatic effects on complex human traits, we focused on the epigenome, a complex molecular phenotype under close genetic control[21–24], and here more specifically on DNA-CpG methylation, measured with the Infinium HumanMethylation450 BeadChip Kit (Illumina 450 K), with DNA derived from blood in two independent samples. We performed an exhaustive search of epistatic

[1]Division of Molecular Neuroscience, Department of Psychology, University of Basel, CH-4055, Basel, Switzerland. [2]Transfaculty Research Platform Molecular and Cognitive Neurosciences, University of Basel, CH-4055, Basel, Switzerland. [3]Psychiatric University Clinics, University of Basel, CH-4055, Basel, Switzerland. [4]Department Biozentrum, Life Sciences Training Facility, University of Basel, CH-4056, Basel, Switzerland. [5]Division of Cognitive Neuroscience, Department of Psychology, University of Basel, CH-4055, Basel, Switzerland. [6]sciCORE, Scientific Computing Center, University of Basel, CH-4056, Basel, Switzerland. [7]SIB - Swiss Institute of Bioinformatics, CH-1015, Lausanne, Switzerland. Correspondence and requests for materials should be addressed to A.M. (email: annette.milnik@unibas.ch)

	Discovery sample		Replication sample	
	All subjects data freeze 2013-08	Selected subjects	All subjects data freeze 2014-04	Selected subjects
Sample size N	1174	533	1935	319
Sex female	59.8%	58.3%	66.1%	69.6%
Blood sampled	63.7%	100%	36.1%	100%
Affymetrix 6.0 data	84.3%	100%	89.9%	100%
Genetic outlier	6.5%	0%	7.8%	0%
Age at main investigation	22 (18–35)	22 (18–35)	23 (18–35)	23 (18–35)
Age at blood sampling	23 (18–36)	23 (18–36)	24 (18–39)	24 (18–36)
Days between main investigation and blood drawing	336 (1–1392)	350 (2–1385)	642 (1–1992)	380 (1–954)
Smoking behavior at main investigation	1.6 (1–5)	1.6 (1–5)	1.8 (1–5)	1.7 (1–5)

Table 1. Sample description. Phenotypic information was collected at the time-point of the main investigation (see Methods). Subjects were later re-invited for an additional blood sampling to investigate e.g. blood-related methylation and expression values. Reported are the numbers from the data freezes used to select subjects for the blood-DNA-methylation study (discovery sample 2013-08; replication sample 2014-04). Quantitative variables are reported as mean (min - max). Smoking behavior was measured on a 5-point Likert-scale ranging from 1 (never) up to 5 (20 cigarettes per day).

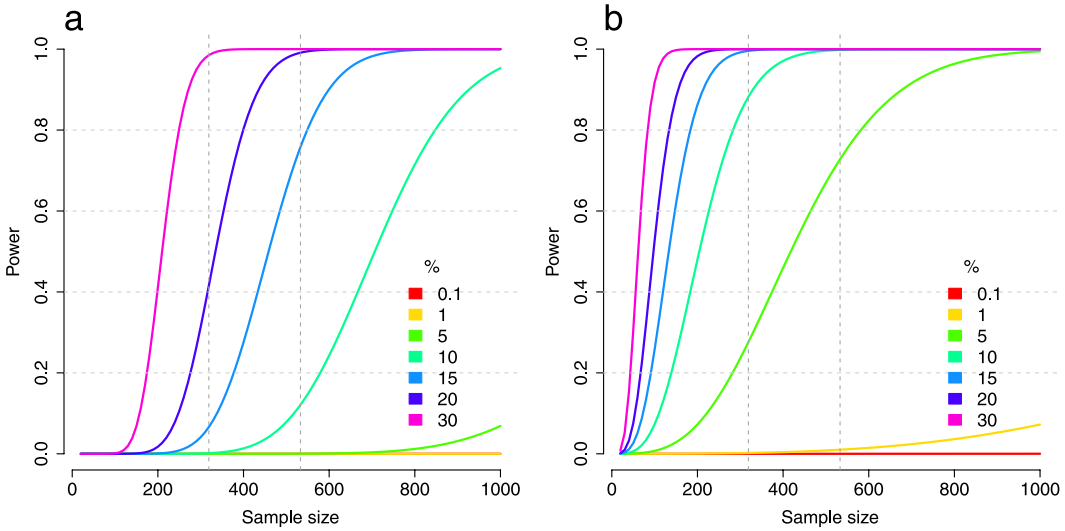

Figure 1. Power analysis exhaustive search for epistatic effects. In **(a)** we adjusted alpha to reach genome-wide and methylome-wide Bonferroni correction (discovery phase, $p = 6.8 \times 10^{-18}$). In **(b)** we adjusted alpha to reach a per-CpG Bonferroni correction threshold (replication phase, $p = 3.8 \times 10^{-6}$). The legends depict the variance that can be explained (in percentage) for different effect sizes ($r_{min} = 0.03, 0.1\%$; $r_{max} = 0.55, 30\%$). The vertical gray bars correspond to a sample size of $N = 533$ (discovery sample) and $N = 319$ (replication sample).

effects by applying several screening and replication steps and subsequent model-confirmation. Additionally, we performed a search for epistatic effects based on SNPs proximal to CpGs, which are exhibiting main effects.

Results

We used data from two independent samples of healthy young adults from the general population (discovery sample $N = 533$, replication sample $N = 319$; see Table 1; DNA was derived from blood, see Methods)[25]. With these sample-sizes we were adequately powered to detect and replicate medium to strong effect sizes in an exhaustive search (see Fig. 1). Both samples have a comparable genetic (see Supplementary Fig. S1) and phenotypic background (see Table 1). SNP-data ($N = 517{,}504$ SNPs) as well as CpG-data ($N = 395{,}431$ CpGs; see Supplementary Figs S2 and S3 for diagnostic plots of the data) was selected to reach high quality metrics in both samples (see Methods).

Discovery phase of the exhaustive search. In the discovery sample, we applied a two-step approach to first identify and then confirm interaction effects. First, based on a pruned set of SNPs ($N = 192{,}955$ SNPs) with sufficient minor allele frequency (see Methods), we performed an exhaustive screening ($N = 7.36 \times 10^{15}$ interaction analyses) with EpiGPU[26]. EpiGPU is optimized for high-speed genome-wide interaction analyses, but

	N unique CpGs	Average *N* hits per CpG	Max *N* hits per CpG	*N* hits in total	Both SNPs in *cis*	One SNP in *cis* and in *trans*	Both SNPs in *trans*
Before replication	13,112	657	46,314	8,608,567	0.03%	0.20%	99.78%
After replication	1,477	3	131	4,816	43.60%	10.36%	46.03%
+Permutation and sign-test	802	3	49	2,262	90.45%	3.98%	5.57%
Per-CpG model approach	174	1	5	239	88.28%	4.18%	7.53%
+Exclusion of LD-block associated effects	47	1	3	55	63.64%	18.18%	18.18%

Table 2. Main results exhaustive search for epistatic effects. The results shown refer to significant interaction effects, depending on the different analytical steps. *cis* is defined as 500 KB around the CpG.

uses a simplified parameterization of the interaction term (see Methods). We filtered SNP-pairs that fulfill basic criteria with respect to the number of groups (all 3×3 genotype combinations existing) and estimated *F*-statistic of their interaction term ($F_{int} > 22$, which was close to Bonferroni correction). Analyses that survived this filtering step ($N = 9.54 \times 10^9$) entered a recalculation phase, in which we reproduced the interaction *F*-statistic with a linear model approach featuring an increased accuracy at the cost of a higher computational burden. Only analyses that survived genome-wide Bonferroni-correction ($p_{int} < 6.8 \times 10^{-18}$, correcting for all 7.36×10^{15} initial interaction analyses) after recalculation and in which the minimal group size was above 3 in both the discovery and replication sample entered the replication phase ($N = 8{,}608{,}567$ analyses; $N = 13{,}112$ unique CpGs; see Table 2). To rule out that the above described two-step procedure is overly conservative, we additionally applied the Bonferroni-correction directly to the *p*-value derived from the original EpiGPU analysis, which resulted in a nearly identical outcome: $N = 8{,}658{,}122$ analyses survived based on $N = 13{,}214$ unique CpGs.

Replication phase of the exhaustive search. Replication was also done in two steps. Firstly, we selected interaction effects that survived a per-CpG Bonferroni-correction in the replication sample (i.e. correcting for at least 13,112 tests; see Methods; $p_{int} < 3.8 \times 10^{-6}$; $N = 4{,}816$ analyses survived). Because of the unequal distribution of subjects in the 9 combined genotype-groups, assignment of few subjects with phenotypes from an extreme end of the distribution to one cell may lead to false-positive findings. Hence, we additionally applied a per-CpG permutation approach to confirm the significance of these findings with an empirical *p*-value (same per-CpG *p*-value threshold with empirical $p_{int} < 3.8 \times 10^{-6}$; 51% of analyses discarded), as well as a sign-test to filter out results that show inconsistent directions of effect between samples (22% of analyses discarded; see Supplementary Figure S4). 2,262 analyses (47%), based on 802 unique CpGs survived both steps (see Table 2; see Methods for a comparison between the above described procedure with a Bonferroni correction for all tests performed).

Per-CpG modeling. Importantly, two SNPs in linkage disequilibrium (LD) with a third SNP that shows a main effect on the trait of interest might, under certain circumstances, mimic an interaction effect that is in fact fully attributable to the main effect[20,27] (see Fig. 2 for one example). Therefore, in the next validation step of epistatic effects, we aimed at simultaneously taking into account main effects as well as interaction effects by generating one comprehensive model of SNP-effects for each single CpG. For this analysis we used the entire SNP-set yielding a higher resolution compared to the smaller SNP-set ($N = 517{,}504$ SNPs instead of $N = 192{,}955$ SNPs). We applied a stepwise-forward regression approach, starting with main effects before adding interaction effects. For $N = 174$ CpGs at least one significant interaction effect was detectable (239 SNP-pairs in total; see Table 2) in both samples, when also taking into account significant main effects of SNPs (see Supplementary Table S1 and Supplementary Fig. S6 for the full models).

Estimation of epistatic effects can be biased for non-independent genomic loci[27,28]. Therefore, we calculated the LD for the SNP-pairs showing interaction effects. We empirically determined the threshold of LD that was unlikely to occur by chance in the discovery sample ($r^2 > 0.021$, $p < 0.001$, see Supplementary Fig. S5). 184 out of the 239 SNP-pairs (77%) showed an $r^2 > 0.021$ in the discovery sample; these pairs were all located in close proximity (mean = 53 KB, min = 0.5 KB, max = 444 KB) to each other on the same chromosome. Hence, we classified these effects as LD-block associated effects[1,29]. After exclusion of LD-block associated effects 47 CpGs (based on 55 SNP-pairs) still showed at least one significant epistatic effect (see Table 2). For the remaining 127 CpGs we identified only LD-block associated effects.

When considering all 174 CpGs, significant main effects explained on average 57% of variance from the CpG signals whereas interaction effects additionally explained on average 8% of variance (see Table 3).

Search for epistatic effects based on SNPs exhibiting main effects. We performed an additional search for interaction effects, based on SNPs in proximity to the CpG (defined as 3.5 MB window, number of SNPs per CpG ranging from 1 to 3,290, mean = 1,223) that show a significant main effect. The 3.5 MB window comprises >95% of all main effects from SNPs on CpGs[25]. By focusing on SNPs being located in proximity to CpGs and exhibiting main effects, this approach was computationally less expensive ($N = 4.84 \times 10^8$ main effects and $N = 7{,}173{,}795$ interaction effects tested) in comparison to the exhaustive search ($N = 7.36 \times 10^{15}$ interaction analyses).

We first applied Bonferroni-correction per CpG in the discovery sample (*p*-value ranging from 0.05 to 1.5×10^{-5}) to identify significant main effects and then tested all significant findings in the replication sample. After replication, $N = 59{,}134$ CpGs showed at least one significant main effect of a SNP in the 3.5 MB window

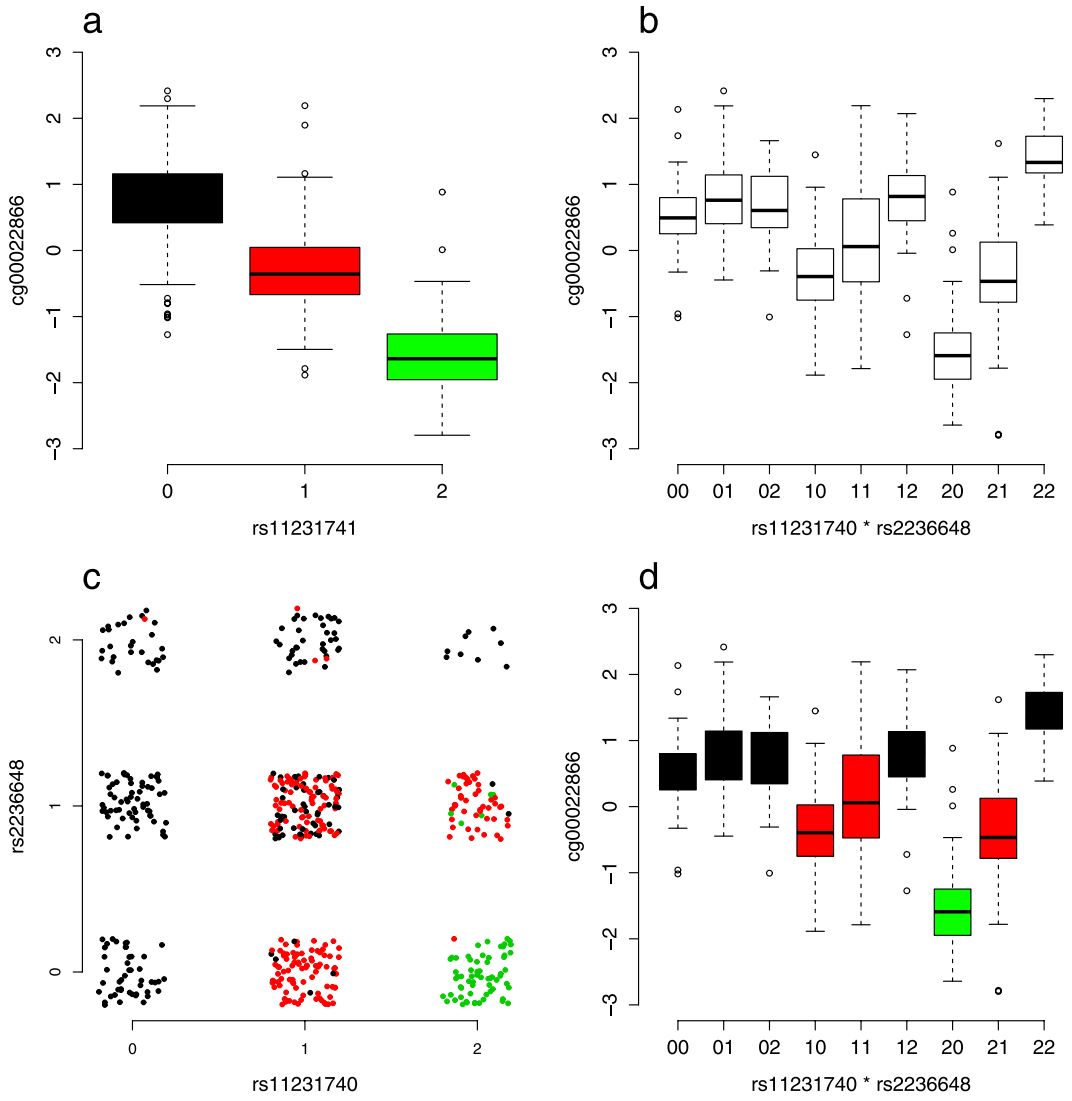

Figure 2. Example of a main effect of a SNP causing a spurious significant interaction effect between two other SNPs. Data is shown for cg00022866 from the discovery sample. **(a)** rs11231741 shows a strong main effect ($p = 4.5 \times 10^{-112}$). This causes a spurious significant interaction (**b** $p = 3.3 \times 10^{-18}$) because rs11231741 is in LD with both interacting SNPs (rs11231740: $r^2 = 0.55$; rs2236648: $r^2 = 0.25$). Of note, the two interacting SNPs show low LD only ($r^2 = 0.024$). Panel **(c)** depicts the dependencies between the 9 SNP-groups build from rs11231740 and rs2236648 and the 3 SNP-groups from rs11231741 (color-coded in black, red and green; a jitter has been added to the data): the 9 SNP-groups of the interacting SNPs mimic the three SNP-groups of the main effect, with 5 of the 9 groups mainly corresponding to the homozygous common allele carrier (black), 3 of the 9 groups mainly corresponding to the heterozygous group (red) and 1 group mainly corresponding to the homozygous rare allele carrier (green). Panel **(d)** shows the same data as in **(b)**, but now with color-coding of the three SNP-groups from rs11231741.

(mean = 9, max = 508). The observation that ~15% of CpGs were associated with a genetic marker is in agreement with previous results[24,30], considering differences in sample-sizes between studies and hence differences in power to detect such effects.

Out of these 59,134 CpGs, $N = 48{,}293$ CpGs showed more than one significant main effect of a SNP. For these CpGs, we identified all significant interaction effects between SNPs that show a significant main effect (see Methods). $N = 24{,}892$ interaction effects based on $N = 3{,}564$ CpGs survived this step (number of interaction effects per CpG: mean = 7, max = 327). Based on these results, we again performed a forward regression approach for each CpG. We included all identified main effects and interaction effects in this analysis, testing main effects first. In the forward regression approach $N = 364$ interaction effects remained significant, based on $N = 281$ unique CpGs (number of interaction effects per CpG: mean = 1, max = 6; see Supplementary Table S2 for the full models). $N = 255$ SNP-pairs (70%) revealed an $r^2 > 0.021$, and were classified as LD-block associated effects. After exclusion of LD-block associated effects, $N = 91$ CpGs showed at least one epistatic effect. The remaining $N = 190$ CpGs showed LD-block associated effects only. We determined the amount of variance that can be explained by

	Discovery sample average variance explained	Replication sample average variance explained
All significant main effects	57.1%	57.8%
- Most-significant main effect	44.9%	45.4%
All significant interaction effects	8.2%	8.3%
- Most-significant LD-block associated effect	4.4%	4.5%
- Most-significant epistatic effects	7.8%	7.4%
All significant main effects and interaction effects	65.2%	66%

Table 3. Exhaustive search average variance explained by main effects and interaction effects. The results are based on the $N = 174$ CpGs that showed at least one significant interaction effect when taking into account also main effects. For only 7 out of 174 CpGs (4%) no significant main effect of a SNP was detectable. 12 out of 174 CpGs (6.9%) showed both, a LD-block based effect as well as an epistatic effect. All significant main effects: average variance explained by all main effects that were kept in the final model. Most-significant main effect: average variance explained by the main effect that exhibited the smallest p-value. All significant interaction effects: average variance explained by all interaction effects that were kept in the final model; these were further separated in LD-block associated effects with SNP-pairs showing an $r^2 > 0.021$, or epistatic effects ($r^2 \leq 0.021$), most-significant corresponds to the effect with the smallest p-value, if more than one of these effects were kept in the final model. All significant main effects and interaction effects: average variance explained by all main effects and interaction effects that were kept in the final model.

single-SNPs, by all significant main effects of SNPs and by full models that include also interaction effects when applying the forward regression approach (see Table 4). The overall variance explained and the variance explained by single-SNP hits systematically increased with increased model complexity.

The detection rate for CpGs exhibiting significant interaction effects was comparable for the exhaustive search ($N = 174$, 0.044% of all CpGs analyzed) and the search based on SNPs that show a main effect ($N = 281$, 0.071% of all CpGs analyzed). Taking both approaches together, we identified a total of $N = 404$ CpGs showing interaction effects (0.1% of all CpGs analyzed), with 51 CpGs being identified in both analyses (see Supplementary Table S3).

Enrichment analyses. The $N = 404$ CpGs showing interaction effects could be assigned to $N = 350$ clusters, when assigning CpGs with an $r > 0.8$ to one cluster (CpGs per cluster: mean = 1.15, max = 5; the maximal base-pair distance per cluster was 4,820). For the following enrichment-analyses (see Table 5), we randomly chose one CpG per cluster. These $N = 350$ CpGs were significantly underrepresented in CpG islands ($p = 1.1 \times 10^{-6}$) and significantly overrepresented in DNase I hypersensitivity sites ($p = 0.025$). Furthermore, these CpGs showed a strong enrichment in significant associations with gene expression ($p = 1.2 \times 10^{-127}$), with the top-hit of the transcripts being located in *cis* of the CpG in all cases except two (see Supplementary Table S3). Gene-set enrichment analysis (see Table 6) on those 350 CpGs showed significant association signals ($p_{FDR} < 0.05$) for 7 KEGG-pathways, with the top-hits being associated with Human papillomavirus (HPV) infection (hsa05165, $p_{fdr} = 0.0014$) and cancer (hsa05200, $p_{fdr} = 0.0036$).

Discussion

Our results demonstrate the existence of higher-order genetic effects on DNA CpG methylation. However, it is important to stress that the absolute number of CpGs that showed replicable higher-order genetic effects was low ($N = 404$ CpG in total, 0.1% of all CpGs). The major fraction of SNP-pairs showing significant interactions were located in *cis* of the CpG, with both SNPs being in LD, which might lead to biased estimates[27,28]. Interestingly, for CpGs exhibiting interaction effects we detected significant enrichment for DNase I hypersensitivity sites paralleled with strong enrichment of significant associations with gene expression. The underrepresentation of these CpG-sites in CpG islands further support their association with transcriptionally active regions. Finally, these CpGs were enriched in pathways related to HPV-infection and cancer.

We stress that the success of the analysis performed herein strongly relies on effect-sizes. With our data we were adequately powered to detect and replicate medium to strong effect sizes in the exhaustive search. However, especially for effects in *trans* we can expect a more complex picture with an accumulation of small effect sizes[31]. Under this scenario, the sample size needed to detect these effects is considerably larger.

Despite the fact that we focused an a complex trait that is biologically very close to the genetic variation[21–23], we did not find strong evidence for the existence of epistatic effects. This result adds to the ongoing debate of the existence and relevance of epistasis[5,6]. We note that the computational effort to approach questions related to epistatic effects is considerably larger in comparison to investigating independent additive models, especially in the context of an exhaustive search. This speaks more in favor of refining the simple additive model before adding another level of complexity by including epistatic effects. Focusing on SNPs that show a significant main effect on the CpG resulted in a similar detection rate in comparison to performing an exhaustive search. This strategy optimizes the ratio between computational burden and the overall detection rate of interaction effects.

Taken together, our results demonstrate the existence of higher-order influences of structural genetic variation on the CpG signal. However, they also show that the impact of these higher-order dependencies is much weaker in comparison to main effects. Interestingly, filtering for CpGs that were under strong and more complex genetic

Per-CpG model in 3.5 MB window	N unique CpGs	Discovery sample Average variance explained by SNPs			Replication sample Average variance explained by SNPs		
		Most-signif. main effect	All signif. main effects	All signif. main effects and interaction effects	Most-signif. main effect	All signif. main effects	All signif. main effects and interaction effects
- CpGs showing at least one significant main effect	59,134	16%	17.7%	17.7%	16%	18.1%	18.1%
- CpGs showing at least two significant main effect	17,938	22.6%	28.2%	28.3%	22.3%	29%	29.1%
- CpGs showing significant interaction effects	281	31.2%	41.9%	46.8%	31.2%	43.3%	49.1%

Table 4. Search for epistatic effects based on SNPs exhibiting main-effects. Average variance explained by main effects of SNPs or interaction effects of SNP-pairs. Results are shown for three different filtering steps, which were based on the number of significant main effects or interaction effects per CpG, identified with a forward-linear regression approach. Most-signif. main effect: average variance explained by the main effect that exhibited the smallest *p*-value. All signif. main effects: average variance explained by all main effects that were kept in the final model. All signif. main effects and interaction effects: average variance explained by all main effects and interaction effects that were kept in the final model. Signif.: significant.

	Expected	Observed	*p*
CpG Island	32.9%	20.6%	1.1×10^{-6}
TFBS	63.3%	66.9%	0.19
DNase I	70.4%	76%	0.025
Gene expression	10.8%	50.9%	1.2×10^{-127}

Table 5. Enrichment analyses. For $N = 404$ CpGs we identified a significant interaction between SNPs. These 404 CpGs could be assigned to $N = 350$ clusters. For each cluster we randomly assigned one CpG as representative. For these 350 CpGs we compared the observed percentage of being located in CpG-dense regions (CpG Island), transcription factor binding sites (TFBS), DNase I hypersensitivity sites (DNase I) or being associated with gene expression against the expected numbers that are based on all remaining CpGs ($N = 395{,}027$), by using Chi^2-tests.

Term	Pathway	*N* genes	*N* hits	$p_{nominal}$	p_{FDR}
hsa05165	Human papillomavirus infection	303	10	1.7×10^{-7}	0.0014
hsa05200	Pathways in cancer	384	10	1.1×10^{-6}	0.0036
hsa05224	Breast cancer	141	7	1.5×10^{-6}	0.0036
hsa01100	Metabolic pathways	1190	14	1.7×10^{-6}	0.0036
hsa04014	Ras signaling pathway	217	7	9.9×10^{-6}	0.017
hsa04151	PI3K-Akt signaling pathway	317	8	1.2×10^{-5}	0.018
hsa05203	Viral carcinogenesis	191	6	2.9×10^{-5}	0.036

Table 6. Results for the gene-set enrichment analysis. Significant gene-sets ($p_{FDR} < 0.05$) are reported.

control increased the power to detect CpGs that are associated with gene expression and biological pathways associated to HPV-infection and cancer.

Methods

Subjects and study design. The subjects included in this blood-DNA-methylation study (see Table 1) represent subsets of two ongoing studies, which were described previously[32,33]. The purpose of the ongoing studies is to identify biological correlates of cognitive performance by using genetics, electroencephalography and imaging techniques in healthy young adults from the general population. Saliva samples and phenotypic information were collected at the time-point of the main investigation. Subjects were later re-invited for additional saliva and blood sampling (see Table 1). Aim of this re-invitation was to additionally collect high-quality DNA from blood e.g. for the study of DNA-methylation and DNA-expression, without assessing further phenotypes. Blood and saliva samples were collected between midday and evening (mean time = 2:30 p.m., range 1:00 p.m.–8.00 p.m.). Hematological analysis, including blood cell counts, was performed with Sysmex pocH-100*i*™ Automated Hematology Analyzer (Sysmex Co, Kobe, Japan).

Subjects were of good general health, did not self-report any neurological or psychiatric illness and did not take any medication (apart from oral contraception) at both time points. The phenotypic data reported here is based on the data freezes that have been used to select the subjects for the blood-DNA-methylation study (discovery sample data freeze 2013-08, $N = 1{,}174$ subjects; replication sample data freeze 2014-04, $N = 1{,}935$ subjects). Subjects were only included in the DNA-methylation study if they had been genotyped previously (Affymetrix

6.0, after QC, see below) and blood had been sampled. For the replication sample, additional requirements were a European genetic background (see below) and a time-distance of less than four years between the main investigation and the blood sampling. About 55% of the subjects from the discovery sample and 28% of the subjects from the replication sample fulfilled these requirements when planning the DNA-methylation study. Individuals were selected randomly from these pools of subjects.

The ethics committee of the Cantons of Basel-Stadt and Basel-Landschaft approved the studies. All participants received general information about the study and gave written informed consent. All methods were performed in accordance with the relevant guidelines and regulations of the participating institutions.

Affymetrix SNP 6.0 based genotyping and imputation. Saliva samples were collected using the Oragene DNA Kit (DNA Genotek, Kanata, Canada). Saliva DNA was extracted from the Oragene DNA Kit using the standard precipitation protocol recommended by the producer. DNA isolated from saliva was investigated with Affymetrix SNP 6.0 array as described in the Genome-Wide Human SNP Nsp/Sty 6.0 User Guide (Affymetrix, Santa Clara, CA USA; see Supplementary text). The mean call-rate per subject was 98.5% (90.1–99.7%).

The genotypic data was projected on the two first PCA components inferred from HapMap reference populations (YRI, CEU and CHB-JPT populations). Outliers were identified using a Bayesian Clustering Algorithm[34]. $N = 35$ subjects out of the discovery sample for which DNA-methylation data was available were identified as outliers and excluded from the statistical analyses.

To reduce the computational burden for the interaction analyses, we used stringent QC-criterion for the exhaustive search for SNP-SNP-interactions: in both samples MAF $> 2\%$, $p_{HWE} > 0.001$, missing rate per SNP $< 1\%$, size of smallest SNP-group ≥ 15. To further eliminate highly redundant information, we additionally applied LD-based-pruning in the discovery sample with the following settings: window-size 50 KB, number of SNPs to shift 5, SNP-SNP $r^2 = 0.95$, resulting in $N = 192{,}955$ SNPs. For the in-depth modeling of additional main effects of SNPs we used a more-comprehensive SNP-set based on the following settings: in both samples MAF $> 2\%$, $p_{HWE} > 0.001$, missing rate per SNP $< 1\%$ ($N = 517{,}504$ SNPs).

To determine a sample-specific threshold for Linkage Disequilibrium, we estimated for 10,000 randomly drawn SNPs from the larger SNP-set the association (r^2) with 10,000 random SNPs located on a different chromosome, based on the data from the discovery sample. We set the LD-threshold to $r^2 > 0.021$, which was very unlikely ($p < 0.001$) to happen by chance between independent SNPs in discovery sample (see Supplementary Fig. S5).

HumanMethylation Infinium 450 K BeadChip based methylation analyses. DNA isolated from peripheral blood was investigated with the Illumina 450 K array (Illumina, Inc., San Diego, CA, U.S.A; see Supplementary text). Postprocessing was done for each sample separately, combining the β-values of the preprocessed data of all batches per sample (see Supplementary text). The β-values were processed step-by-step in order to correct for further influential and putative confounding factors: (1) using logit-transformation (M-value[35], done with the R-package car[36]); (2) z-transformation per plate (correcting for plate and batch effects); (3) regressing out the first 8 (discovery sample) or 7 (replication sample) axes of a principal component analysis (PCA, done with the R-package pcaMethods[37]). The PCA was based on CpGs with no missing values (>95% of the included CpGs). The PCA-based approach corrected for technical biases as well as for part of the variability induced by blood cell composition[25] (4) regressing out the effects of sex and age; (5) regressing out the effects of variants in the 50mer probe sequence, if the total variance explained by these variants exceeded 0.1% (see below). The accepted missing rate per CpG was set to <1% in both samples. We further excluded cross-hybridizing probes and polymorphic CpGs sites[38,39] ($N_{max} = 63{,}974$). Only CpGs surviving all filtering steps in both samples were used for the downstream analyses ($N = 395{,}431$).

Correction for genetic variants in the 50mer probe sequence. We performed imputation of the genetic data: Prior to autosome-wide genotype imputation, haplotype estimation was performed using SHAPEITv2[40], allowing a per individual and a per SNP missing rate for observed markers of max. 5%. After pre-phasing, genotype imputation was performed using IMPUTE v2.3.0, which imputes missing genotypes using a multi-population reference panel[41,42]. The integrated variant callset of 1,092 individuals from the 1000 Genomes Project (release v3 in NCBI build 37/hg19 coordinates, March 2012) served as panel data (http://mathgen.stats.ox.ac.uk/impute/ALL_1000G_phase1integrated_v3_impute_macGT1.tgz).

Based on the genomic location we filtered for all 50mer probe-sequences that comprises imputed variants. As a very basic QC, we excluded imputed variants with a minor allele frequency (MAF) <0.03% in our population (based on information of $N = 3{,}166$ subjects). For each 50mer probe sequence containing at least one variant, we build a linear model with the probability information of all imputed genotypes of this probe sequence as independent variables, and the CpG signal as dependent variable. If the overall explained variance of this model exceeded 0.1%, we used the residuals of this model as 50mer-corrected CpG signal. This procedure was done independently for the discovery sample ($N = 533$ subjects after outlier exclusion) and the replication sample ($N = 319$ subjects). Out of the 395,431 CpGs, $N = 121{,}868$ were corrected for 50mer variants in at least one of the two studies (discovery sample $N = 98{,}532$; replication sample $N = 103{,}462$; overlap $N = 80{,}126$).

Exhaustive search for epistatic effects – Discovery phase. In the discovery sample we performed an exhaustive genome-wide search for two-locus (SNP-SNP) epistatic effects ($N = 192{,}955$ SNPs; $N = 1.85 \times 10^{10}$ SNP-pairs) on single CpG methylation levels ($N = 395{,}431$ CpGs), resulting in 7.36×10^{15} tests ($N_{CpGs} * N_{snps} * (N_{snps} - 1)/2$) to calculate. Accordingly, the Bonferroni-corrected threshold (alpha 5%) was set to $p < 6.8 \times 10^{-18}$.

The statistical analysis in the discovery phase was based on a simplified interaction analysis strategy that corresponds to an analysis of variance based on 9 genotype groups as fixed effects[26]. The 8 degrees of freedom (df) test determines the joint genetic effects at two loci (combination of main and interaction effects). For the interaction analyses only, the additive and dominance effects at each locus were subtracted from the mean effects of each pairwise genotype. This 4 df parameterization is an approximation to a true interaction test[1,26]. Independence between the SNPs is a prerequisite for an accurate 4 df test, which is often not fulfilled, especially for SNPs in proximity to each other. Additionally, the accuracy of the interaction approximation decreases if there is a large (additive or dominant) main effect of a SNP. To minimize the bias of the simplified interaction analysis on the results, we recalculated all results with an F-value > 22 (4 df; $p_{approx} < 8.0 \times 10^{-17}$) of the simplified EpiGPU interaction analysis strategy with a linear regression in combination with an ANOVA-approach. Only analyses that remained significant after Bonferroni-correction entered the next analytical step. In most of these cases (99.89%) the simplified *F*-value was larger than the *F*-value based on recalculation, indicating an over-estimation of the true effects when using the simplified *F*-value only.

Exhaustive search for epistatic effects – Replication phase. In the replication sample we applied a per-CpG Bonferroni-corrected *p*-value threshold. As baseline value for the multiple testing correction we used the number of $N = 13'112$ unique CpGs ($p < 3.8 \times 10^{-6}$). In case a CpG showed multiple interaction signals, we applied a more stringent correction by additionally accounting for the number of interaction hits per CpG ($p_{int} <$ [0.05/(13′112 + number of interaction hits per CpG]). This procedure resulted in a per-CpG *p*-value threshold ranging from 3.8×10^{-6} (one SNP-pair only) up to 8.4×10^{-7} (46′314 SNP-pairs). To additionally derive an empirical *p*-value we performed on average 1.70×10^{7} permutations of the CpG-signal and recalculated the interaction analysis with the permuted signals to derive an empirical *F*-distribution, for each CpG separately. We performed a sign-test by comparing the average methylation level of the 9 SNP-groups between the two samples with Pearson correlations *r* (see Supplementary Fig. S4). Analyses had to pass all three filters (per-CpG Bonferroni correction based on *p*-value derived from *F*-distribution and from empirical *F*-distribution and sign-test $r > 0.85$).

We tested the stringency of the per-CpG empirical *p*-value ($p_{emp} < 3.8 \times 10^{-6}$) in comparison to a global Bonferroni-correction for all tests (alpha 5%, $p_{bonf} < 5.8 \times 10^{-9}$, correcting for 8′608′567 tests, with *p*-values derived from the F-distribution). The empirical *p*-value was as at least as stringent (49% of analysis results survived) as Bonferroni-correction for all tests (52% of all analysis results survived) and had a higher concordance rate with the sign-test: only 4% of analysis surviving the empirical *p*-value failed the sign-test, in comparison to 10% when applying Bonferroni-correction for all tests.

Per CpG modeling. For this analysis we used the entire SNP-set ($N = 517'504$ SNPs instead of $N = 192'955$ SNPs), allowing for a better resolution. We applied a stepwise-forward regression approach, starting with main-effects. For each CpG, we first tested main effects of SNPs in a 500 KB window around the CpG and around the SNPs showing an interaction effect, as well as all SNPs that showed a significant main effect on a genome-wide scale (Bonferroni correction $p_{main\text{-}effect} < 9.7 \times 10^{-8}$). We then also included interaction effects by testing all already identified interacting SNP-pairs as well as all possible SNP-pairs for SNPs that were in LD ($r^2 > 0.021$ in the discovery sample) with at least one of the identified interacting SNPs. SNPs and SNP-pairs entering the forward regression model were sorted by their main effect *p*-value and interaction effect *p*-value of the discovery sample. We kept SNPs in the model based on their main effect and SNP-pairs based on their interaction effect, if their *p*-value in the forward-regression analysis was smaller than $p =$ [0.05/(number of main-effects + number of interaction-effects tested per CpG)] in both the discovery and the replication sample.

Search for epistatic effects based on SNPs exhibiting main effects. For each CpG we restricted the analysis to SNPs located within a ± 3.5 MB window around the CpG, using the larger SNP-set ($N = 517{,}504$ SNPs). For these SNPs we evaluated linear models with ANOVAs by using a 2-df parameterization of the SNP's main effect on the CpG signal. We searched for main effects of SNPs that survived a per-CpG Bonferroni correction accepting an alpha error of 5% per CpG, by correcting for the number of main effects tested per CpG in the discovery sample. SNP-effects surviving this analysis were tested in the replication sample, again by using a per-CpG Bonferroni correction accepting an alpha error of 5%, correcting for the number of main effects tested in the replication sample. For CpGs showing at least two significant main effects, we tested for significant interaction effects in the discovery and replication sample, restricting the analyses to interaction effects in which the minimal group size was larger than 3 in both, the discovery and replication sample. We applied a per-CpG Bonferroni correction (alpha error 5%, correcting for the number of interaction effects tested) in both samples to identify significant interactions. Next, for each CpG we run a stepwise forward regression approach that included all significant main effects and interaction effects. We sorted all SNPs by their main effect and all SNP-pairs by their interaction effect *p*-value of the discovery sample. Main effects entered the forward regression analysis first. In the model we kept SNPs based on their main effect and SNP-pairs based on their interaction effect, if their *p*-value in the forward-regression analysis was smaller than $p =$ [0.05/(number of main effects + number of interaction effects tested per CpG)] in both the discovery and the replication sample.

Affymetrix HTA 2.0 array transcriptome analysis. Blood samples were collected using PAXgene Blood RNA Tubes (PreAnalytix Qiagen/BD, Switzerland). Expression profiles were measured with the Affymetrix GeneChip Human Transcriptome Array 2.0 (see Supplementary text). Individual expression values of each transcript were adjusted for age, sex and 23 components of a PCA by using linear regression models, based on data from $N = 416$ unrelated subjects. The PCA was derived from the expression data, 23 components of the PCA were chosen to optimize the signal-to-noise ratio for association analyses with genetic marker.

Enrichment analyses. We used the genomic hg19 database (genome-mysql.cse.ucsc.edu; accessed 2016-08) to retrieve data about the location of CpG Islands (table cpgIslandExt), DNase I hypersensitivity sites (table wgEncodeR- egDnaseClusteredV3) and transcription factor binding sites (table wgEncodeRegTfbsClusteredV3). For $N = 408$ of our subjects, phenotypic data (data freeze 2015-09), transcriptomic data ($N = 63{,}280$ transcripts; see above) as well as DNA-methylation data ($N = 395{,}431$ CpGs) was available. For each CpG we calculated a genome-wide association analysis (Pearson's correlation coefficient) with the transcriptomic data ($N = 2.50 \times 10^{10}$ analyses). We identified significant CpG-transcriptome associations on a genome-wide scale (alpha = 5%, Bonferroni correction, $p_{bonf} < 2.0 \times 10^{-12}$) and in *cis* (alpha = 5%, $p_{cis} = 0.05/N_{local\_transcripts}$; $N_{local\_transcripts}$ is the number of transcripts per CpG within a 500 KB-window, mean = 40, min = 0, max = 186).

For CpGs showing significant interactions, we estimated the similarity between CpGs with Pearson's correlations. CpGs with an $r > 0.8$ were assigned to one cluster. For each cluster we randomly chose one CpG before performing the enrichment analyses. Each CpG was classified as being located within a CpG Island, a DNase I hypersensitivity site or a transcription factor binding site and whether it was significantly associated with a transcript. We compared the observed frequencies from CpGs that show an interaction effect against the expected frequency from all other CpGs by using Chi^2-tests.

We performed a gene ontology (GO) enrichment analysis using the 'gometh'-function from the R-package missMethyl[43]. The algorithm corrects for the varying numbers of CpGs that can be mapped to single genes. We used both, the GO-database as well as the KEGG-database provided in the package missMethyl, restricting the analysis to pathways with at least 10 members ($N = 8{,}596$ in total, GO-database $N = 8{,}288$ out of 21,671 pathways; KEGG-database $N = 308$ out of 320 pathways). We applied FDR-correction based on the total number of $N = 8{,}596$ pathways included in the analysis.

Software. If not mentioned differently, analyses were conducted in R (version: 2.15.1 and higher, R Development Core Team, 2012).

Data availability. The data that support the findings of this study are available from the corresponding author upon request.

References

1. Wei, W.-H., Hemani, G. & Haley, C. S. Detecting epistasis in human complex traits. *Nat. Rev. Genet.* **15**, 722–733 (2014).
2. Manolio, T. A. *et al*. Finding the missing heritability of complex diseases. *Nature* **461**, 747–53 (2009).
3. Hemani, G., Knott, S. & Haley, C. An evolutionary perspective on epistasis and the missing heritability. *PLoS Genet.* **9**, e1003295 (2013).
4. Hirschhorn, J. N., Lohmueller, K., Byrne, E. & Hirschhorn, K. A comprehensive review of genetic association studies. *Genet. Med.* **4**, 45–61 (2002).
5. Crow, J. F. On epistasis: why it is unimportant in polygenic directional selection. *Philos. Trans. R. Soc. London B Biol. Sci.* **365**, 1241–1244 (2010).
6. Hill, W. G., Goddard, M. E. & Visscher, P. M. Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. *PLoS Genet.* **4**, e1000008 (2008).
7. Mackay, T. F. C. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. *Nat. Rev. Genet.* **15**, 22–33 (2013).
8. Cordell, H. J. Detecting gene-gene interactions that underlie human diseases. *Nat. Rev. Genet.* **10**, 392–404 (2009).
9. Blalock, H. M. J. The Identification Problem and Theory Building: The Case of Status Inconsistency. *Am. Sociol. Rev.* **31**, 52–61 (1966).
10. Kreft, I. G. G., Kreft, I. & de Leeuw, J. *Introducing Multilevel Modeling*. (SAGE Publications, 1998).
11. Greenland, S. Basic problems in interaction assessment. *Environ. Health Perspect*. 59–66 (1993).
12. Visscher, P. M. Sizing up human height variation. *Nat Genet* **40**, 489–490 (2008).
13. Sullivan, P. F. The psychiatric GWAS consortium: big science comes to psychiatry. *Neuron* **68**, 182–186 (2010).
14. Ripke, S. *et al*. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
15. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
16. van der Sijde, M. R., Ng, A. & Fu, J. Systems genetics: From GWAS to disease pathways. *Biochim. Biophys. Acta* **1842**, 1903–1909 (2014).
17. Civelek, M. & Lusis, A. J. Systems genetics approaches to understand complex traits. *Nat. Rev. Genet.* **15**, 34–48 (2013).
18. Brown, A. A. *et al*. Genetic interactions affecting human gene expression identified by variance association mapping. *Elife* **3**, e01381 (2014).
19. Hemani, G. *et al*. Detection and replication of epistasis influencing transcription in humans. *Nature* **508**, 249–253 (2014).
20. Wood, A. R. *et al*. Another explanation for apparent epistasis. *Nature* **514**, E3–5 (2014).
21. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500 (2016).
22. Ritchie, M. D., Holzinger, E. R., Li, R., Pendergrass, S. A. & Kim, D. Methods of integrating data to uncover genotype–phenotype interactions. *Nat. Rev. Genet.* **16**, 85–97 (2015).
23. Bell, J. T. & Spector, T. D. DNA methylation studies using twins: what are they telling us? *Genome Biol.* **13**, 172 (2012).
24. Cheung, W. A. *et al*. Functional variation in allelic methylomes underscores a strong genetic contribution and reveals novel epigenetic alterations in the human epigenome. *Genome Biol.* **18**, 50 (2017).
25. Milnik, A. *et al*. Common epigenetic variation in a European population of mentally healthy young adults. *J. Psychiatr. Res.* **83**, 260–268 (2016).
26. Hemani, G., Theocharidis, A., Wei, W. & Haley, C. EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinformatics* **27**, 1462–1465 (2011).
27. Wood, A. R. *et al*. Allelic heterogeneity and more detailed analyses of known loci explain additional phenotypic variation and reveal complex patterns of association. *Hum. Mol. Genet.* **20**, 4082–4092 (2011).
28. Ueki, M. & Cordell, H. J. Improved Statistics for Genome-Wide Interaction Analysis. *PLoS Genet.* **8**, e1002625 (2012).
29. Haig, D. Does heritability hide in epistasis between linked SNPs? *Eur. J. Hum. Genet.* **19**, 123 (2011).
30. Shi, J. *et al*. Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. *Nat. Commun.* **5**, 3365 (2014).
31. Gaunt, T. R. *et al*. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.* **17**, 61 (2016).

32. Heck, A. *et al*. Converging genetic and functional brain imaging evidence links neuronal excitability to working memory, psychiatric disease, and brain activity. *Neuron* **81**, 1203–1213 (2014).
33. Spalek, K. *et al*. Sex-dependent dissociation between emotional appraisal and memory: a large-scale behavioral and fMRI study. *J. Neurosci.* **35**, 920–935 (2015).
34. Bellenguez, C., Strange, A., Freeman, C., Donnelly, P. & Spencer, C. C. A. A robust clustering algorithm for identifying problematic samples in genome-wide association studies. *Bioinformatics* **28**, 134–135 (2012).
35. Du, P. *et al*. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).
36. Fox, J. & Weisberg, S. *An R Companion to Applied Regression* (Sage, 2011).
37. Stacklies, W., Redestig, H., Scholz, M., Walther, D. & Selbig, J. pcaMethods–a Bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **23**, 1164–1167 (2007).
38. Chen, Y. *et al*. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8**, 203–209 (2013).
39. Price, M. E. *et al*. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* **6**, 4 (2013).
40. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
41. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
42. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 (Bethesda).* **1**, 457–470 (2011).
43. Phipson, B., Maksimovic, J. & Oshlack, A. missMethyl: an R package for analyzing data from Illumina?s HumanMethylation450 platform. *Bioinformatics* **11**, btv560 (2015).

Acknowledgements

The exhaustive search for epistatic effects based on EpiGPU was performed at the Swiss National Supercomputing Centre (CSCS) under project ID s521. Consecutive analyses were performed at sciCORE (http://scicore.unibas.ch/), the scientific computing center of the University of Basel, with support by the SIB - Swiss Institute of Bioinformatics. This work was funded by the Swiss National Science Foundation (individual grant 320030_159740 to D.Q.) and was supported by a grant from the Swiss National Supercomputing Centre (CSCS) under project ID s521.

Author Contributions

T.E. performed experiments, analyzed data and wrote the paper. A.M. and V.V. designed the study, performed experiments, analyzed data, and wrote the paper. A.P., C.V. and D.Q. designed the study, contributed to the results interpretation and wrote the paper. T.S., D.C., V.F., A.H. contributed to the data analysis or results interpretation and wrote the paper. T.S., M.J., R.C. contributed to the data analysis in the high-performance computing environments.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-13256-9.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

5. Discussion

Large multidimensional datasets are increasingly becoming available for psychological research. In this thesis, I have pointed out how complex analyses of large multidimensional datasets may enable insights into the complex traits investigated in psychology. I have also proposed requirements regarding informatics infrastructure and data management for conducting such analyses comprehensibly and reproducibly. I have furthermore underlined the importance of counteracting false-positive results, e.g. by applying dimensionality reduction as well as conducting apt validation procedures, like replication and cross-validation. Finally, I have outlined how using elaborate data visualizations, and integrating results or data from past studies alleviate the interpretation of findings. Following these suggestions, we have conducted two large-scale studies on highly multidimensional datasets from distinct modalities.

In the first study (Egli et al., 2018), we reduced the dimensionality of working memory brain activation from ~1'000 × 50'000 data points per subject to six data points per subject using ICA. This data reduction massively increased the statistical power and reduced the false-positive rate for associating the task brain activation with other measurements, such as task performances and structural brain characteristics. Importantly, association analyses on the independent components showed that the data reduction retained (or even slightly refined) the information from the brain activation that is relevant for task performances. The statistically independent brain activation networks estimated by the ICA correspond well to the common assumption of intrinsic functional brain networks (Cole, Ito, Bassett, & Schultz, 2016). ICA (respectively the underlying eigenvalue decomposition) is computationally expensive when applied to large datasets; using a multi-level approach that computed the ICA on

the group-level minimized the computational burden in our analysis. We further assessed whether it is possible to generalize the estimated networks beyond our study sample by comparing them with meta-analytic results from 11'406 fMRI studies (Yarkoni et al., 2011). The comparisons suggested that the spatial characteristics of our main finding, a parietally-centered network associated with working memory performance, had been observed across a large number of prior studies and thus represent a generalizable brain network. We furthermore demonstrated the robustness and validity of our results using resampling and cross-validations. We publicly share our brain networks for application in further studies. Future studies can therefore use the reduced data or adopt our approach to efficiently combine brain activation with other highly dimensional data, e.g. genetic and epigenetic measurements. We have applied the dimensionality reduction to summary statistics of brain activation on the group level (i.e. across all subjects) in order to reduce computational costs. The observed working memory brain networks should be further validated by conducting ICA decompositions on the level of individual subjects, as e.g. described in Erhardt et al. (2011), and comparing the results to our approach.

Our second study (Egli et al., 2017) assessed to which extent DNA methylation is affected by two-locus epistatic effects of SNPs. The study aimed at exhaustively searching for epistatic effects across the human genome. Therefore, reducing the dimensionality of the data prior to the analysis would not have been purposeful in this case. The exhaustive search resulted in well above 10^{15} (a quadrillion) computations and as a result was computationally extremely intensive with regards to speed and time. We therefore conducted the exhaustive analyses on a supercomputer at the Swiss National Supercomputing Centre. Because the exceptionally large number of

computed hypothesis tests likely led to numerous false-positive results, we firstly applied stringent Bonferroni corrections, and secondly discarded results that did not fully replicate in an independent sample. After taking into account additional sources of biases, such as main effects of single SNPs and interaction effects of SNPs in LD that could lead to spurious interaction effects, we identified $N = 404$ CpG-sites that were robustly affected by pairwise SNP-SNP interactions. Of note, when we additionally calculated interaction analyses only across the SNPs showing a main effect onto the CpGs, we attained a similar detection rate for interaction effects as compared to the exhaustive analysis. This cost-efficient approach could thus serve as an approximation to the exhaustive analysis in future studies. Our additional gene-set enrichment analyses suggested that the CpG-sites affected by epistasis were implicated in HPV-infection and cancer. The identified CpG-sites were furthermore more strongly associated with gene expression than expected at random. Because our interaction analysis only had sufficient statistical power for detecting medium to strong effect sizes, we cannot rule out a higher number of interactions with smaller effect sizes. Considering on one hand the substantial computational effort and on the other hand the small number of medium to large effects found in our study, our results indicate that using simple additive models without adding SNP-SNP interaction terms should largely suffice for investigating medium to strong genotypic effects onto DNA methylation. Further epistasis analyses with larger samples are required for assessing SNP-SNP interactions onto DNA methylation with smaller effect sizes.

Implementing the approaches described in this thesis allowed us to transform the very large and multidimensional datasets of two studies into concise and interpretable information: we identified independent networks of brain activation

associated with working memory, and showed that DNA methylation-sites are not strongly affected by epistatic effects of genetic markers. The absence of strong epistatic effects on DNA methylation indicates that analyses on DNA methylation can largely focus on additive effects. Alternatively, DNA methylation could be investigated using multivariate approaches, similar to our approach on neuroimaging data in (Egli et al., 2018). Importantly, Freytag et al. (2017) have successfully used dimensionality reductions for associating DNA methylation with cortical thickness and memory performance.

Of note, the datasets investigated in our studies were large and complex when compared to the smaller datasets often analyzed in psychological research (Chen & Wojcik, 2016; Cheung & Jak, 2016), yet relatively small as compared to big data analyses in other scientific fields such as astronomy (Burns et al., 2014), or in tech industry (Chen et al., 2013). Further incorporating techniques and knowledge from such fields will prepare psychological research for investigating even larger and more complex datasets in the future. This will be especially important in light of the decreasing costs and rising efficiency in data acquisition, which will increasingly lead toward "large *n*, really large *p*" problems in science (Spiegelhalter, 2014). Exploiting "new" and relatively cost-efficient data resources like electronic health records, the behavior of individuals in video games, activity on the internet and social media, or data submitted from devices of daily use ("internet of things") will further increase the volume and complexity of available research data in psychology and psychiatry (Monteith et al., 2015).

Current research (in psychology and other fields) suffers from low reproducibility (Munafò et al., 2017; Open Science Collaboration, 2015; Poldrack et al.,

2017). Improving the reproducibility of future analyses should therefore be a decisive objective in psychological research (Szucs & Ioannidis, 2017), especially in view of the increasing complexity of investigated data. The methodological and statistical precautions described in this thesis may serve as a basis for reproducible and comprehensible research.

This thesis has demonstrated that conducting large-scale analyses of complex datasets requires a broad set of methodologies and skills. Some of them exceed the current training of psychological scientists (Chen & Wojcik, 2016; Cheung & Jak, 2016) and need to be further integrated in educational strategies from undergraduate through post-doctorate levels in psychology (Akil et al., 2011). Rigorous statistical and methodological training is indispensable (Munafò et al., 2017; Schwartz et al., 2016) and has to make up for fast advances in statistical methods (Cheung & Jak, 2016). In order to exploit large multidimensional datasets, psychological scientists should broaden their statistical versatility to include "newer" methods like machine learning techniques (Yarkoni & Westfall, 2017). It will moreover be critical to assure proper training in the management of research data (Kleppner & Sharp, 2009; Wilson et al., 2017), as well as in programming languages and basic software development practices (Wilson et al., 2014). Notably, psychological scientists working with large complex datasets will not require expert-level abilities in all the methods and skills described in this thesis. A wealth of software applications (e.g. R-packages or python libraries) written by experts facilitates applying elaborate methods without acquiring full expertise (Chen & Wojcik, 2016).

In conclusion, I have proposed approaches that can serve as a basis for investigating large multidimensional datasets in psychology. On this basis, future

analyses can be developed that investigate even larger or more complex datasets and use even more elaborate methods, for instance advanced machine learning techniques. Despite the relevance of large-scale analyses on multidimensional datasets for psychology, it is and will continue to be indispensable for certain research questions to investigate smaller and less complex datasets from well-designed, randomized, controlled experiments – given that they are sufficiently powered (Ioannidis & Khoury, 2013; Monteith et al., 2015). The analysis of large and complex datasets will not replace, but complement such investigations. By integrating a broad spectrum of information, analyses of large multidimensional datasets may eventually help to elucidate the very complex mechanisms underlying mental health and disease.

6. References

Akil, H., Martone, M. E., & van Essen, D. C. (2011). Challenges and opportunities in mining neuroscience data. *Science*, *331*(6018), 708–713. https://doi.org/10.1126/science.1199305

Beaulieu, C. (2002). The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR in Biomedicine*, *15*(7–8), 435–455. https://doi.org/10.1002/nbm.782

Bedi, G., Carrillo, F., Cecchi, G. A., Slezak, D. F., Sigman, M., Mota, N. B., Ribeiro, S., Javitt, D. C., Copelli, M., & Corcoran, C. M. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. *Npj Schizophrenia*, *1*, 15030. https://doi.org/10.1038/npjschz.2015.30

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*. https://doi.org/10.2307/2346101

Berman, F. (2008). Got data? A guide to data preservation in the information age. *Communications of the Association for Computing Machinery*, *51*(12), 50–56. https://doi.org/10.1145/1409360.1409376 Tools

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., Delano, D., Zhang, L., Schroth, G. P., Gunderson, K. L., Fan, J.-B., & Shen, R. (2011). High density DNA methylation array with single CpG site resolution. *Genomics*, *98*(4), 288–295. https://doi.org/10.1016/j.ygeno.2011.07.007

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes and Development*, *16*, 6–21. https://doi.org/10.1101/gad.947102.6

Bird, A. (2007). Perceptions of epigenetics. *Nature*, *447*, 396–398. https://doi.org/10.1038/nature05913

Blischak, J. D., Davenport, E. R., & Wilson, G. (2016). A quick introduction to version control with Git and GitHub. *PLoS Computational Biology*, *12*(1), e1004668. https://doi.org/10.1371/journal.pcbi.1004668

Bogdan, R., Salmeron, B. J., Carey, C. E., Agrawal, A., Calhoun, V. D., Garavan, H., Hariri, A. R., Heinz, A., Hill, M. N., Holmes, A., Kalin, N. H., & Goldman, D. (2017). Imaging genetics and genomics in psychiatry: a critical review of progress and potential. *Biological Psychiatry*, *82*(3), 165–175. https://doi.org/10.1016/j.biopsych.2016.12.030

Boland, M. R., Hripcsak, G., Shen, Y., Chung, W. K., & Weng, C. (2017). Defining a comprehensive verotype using electronic health records for personalized medicine. *Journal of the American Medical Informatics Association*, *20*, e232–e238. https://doi.org/10.1136/amiajnl-2013-001932

Bouchard, K. E., Aimone, J. B., Chun, M., Dean, T., Denker, M., Diesmann, M., Donofrio, D. D., Frank, L. M., Kasthuri, N., Koch, C., Ruebel, O., Simon, H. D., Sommer, F. T., & Prabhat. (2016). High-performance computing in neuroscience for data-driven discovery, integration, and dissemination. *Neuron*, *92*(3), 628–631. https://doi.org/10.1016/j.neuron.2016.10.035

Burns, R., Vogelstein, J. T., & Szalay, A. S. (2014). From cosmos to connectomes: the evolution of data-intensive science. *Neuron*, *83*(6), 1249–1252. https://doi.org/10.1016/j.neuron.2014.08.045

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–76. https://doi.org/10.1038/nrn3475

Cao, J., & Zhang, S. (2014). Multiple comparison procedures. *JAMA*, *312*(5), 543–544.

https://doi.org/10.1001/jama.2014.9440

Chen, E. E., & Wojcik, S. P. (2016). A practical guide to big data research in psychology. *Psychological Methods*, *21*(4), 458–474. https://doi.org/10.1037/met0000111

Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: a data management perspective. *Frontiers of Computer Science*, *7*(2), 157–164. https://doi.org/10.1007/s11704-013-3903-7

Chen, M., Mao, S., Zhang, Y., & Leung, V. C. M. (2014). *Big data: related technologies, challenges and future prospects*. (S. Zdonik, P. Ning, S. Shekhar, J. Katz, X. Wu, L. C. Jain, D. Padua, X. Shen, B. Furth, V. S. Subrahmanian, M. Hebert, K. Ikeuchi, B. Siciliano, & S. Jajodia, Eds.). Heidelberg: Springer International Publishing. https://doi.org/10.1007/978-3-319-06245-7

Cheung, M. W.-L., & Jak, S. (2016). Analyzing big data in psychology: a split / analyze / meta-analyze approach. *Frontiers in Psychology*, *7*, 738. https://doi.org/10.3389/fpsyg.2016.00738

Cole, M. W., Ito, T., Bassett, D. S., & Schultz, D. H. (2016). Activity flow over resting-state networks shapes cognitive task activations. *Nature Neuroscience*, *19*(12), 1718–1726. https://doi.org/10.1038/nn.4406

Corvin, A., Craddock, N., & Sullivan, P. F. (2010). Genome-wide association studies: a primer. *Psychological Medicine*, *40*(7), 1063–1077. https://doi.org/10.1017/S0033291709991723

Debette, S., Verbaas, C. A. I., Bressler, J., Schuur, M., Smith, A., Bis, J. C., Davies, G., Wolf, C., Gudnason, V., Chibnik, L. B., Yang, Q., DeStefano, A. L., de Quervain, D. J.-F., Srikanth, V., Lahti, J., Grabe, H. J., Smith, J. A., Priebe, L., Yu, L., Karbalai, N., Hayward, C., Wilson, J. F., Campbell, H., Petrovic, K., Fornage, M., Chauhan, G., Yeo, R., Boxall, R., Becker, J., Stegle, O., Mather, K. A., Chouraki, V., Sun, Q., Rose, L. M., Resnick, S.,

Oldmeadow, C., Kirin, M., Wright, A. F., Jonsdottir, M. K., Au, R., Becker, A., Amin, N., Nalls, M. A., Turner, S. T., Kardia, S. L. R., Oostra, B., Windham, G., Coker, L. H., Zhao, W., Knopman, D. S., Heiss, G., Griswold, M. E., Gottesman, R. F., Vitart, V., Hastie, N. D., Zgaga, L., Rudan, I., Polasek, O., Holliday, E. G., Schofield, P., Choi, S. H., Tanaka, T., An, Y., Perry, R. T., Kennedy, R. E., Sale, M. M., Wang, J., Wadley, V. G., Liewald, D. C., Ridker, P. M., Gow, A. J., Pattie, A., Starr, J. M., Porteous, D., Liu, X., Thomson, R., Armstrong, N. J., Eiriksdottir, G., Assareh, A. A., Kochan, N. A., Widen, E., Palotie, A., Hsieh, Y.-C., Eriksson, J. G., Vogler, C., van Swieten, J. C., Shulman, J. M., Beiser, A., Rotter, J., Schmidt, C. O., Hoffmann, W., Nöthen, M. M., Ferrucci, L., Attia, J., Uitterlinden, A. G., Amouyel, P., Dartigues, J.-F., Amieva, H., Räikkönen, K., Garcia, M., Wolf, P. A., Hofman, A., Longstreth Jr., W. T., Psaty, B. M., Boerwinkle, E., DeJager, P. L., Sachdev, P. S., Schmidt, R., Breteler, M. M. B., Teumer, A., Lopez, O. L., Cichon, S., Chasman, D. I., Grodstein, F., Müller-Myhsok, B., Tzourio, C., Papassotiropoulos, A., Bennett, D. A., Ikram, M. A., Deary, I. J., van Duijn, C. M., Launer, L., Fitzpatrick, A. L., Seshadri, S., & Mosley Jr., T. H. (2015). Genome-wide studies of verbal declarative memory in nondemented older people: the cohorts for heart and aging research in Genomic Epidemiology Consortium. *Biological Psychiatry*, *77*(8), 749–763. https://doi.org/10.1016/j.biopsych.2014.08.027

Deepayan, S. (2008). *Lattice: multivariate data visualization with R*. New York, NY: Springer. Retrieved from http://lmdvr.r-forge.r-project.org

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Egli, T., Coynel, D., Spalek, K., Fastenrath, M., Freytag, V., Heck, A., Loos, E., Auschra, B., Papassotiropoulos, A., de Quervain, D. J.-F., & Milnik, A. (2018). Identification of two distinct working memory-related brain networks in healthy young adults. *eNeuro*, *In press*.

Egli, T., Vukojevic, V., Sengstag, T., Jacquot, M., Cabezón, R., Coynel, D., Freytag, V., Heck, A., Vogler, C., de Quervain, D. J.-F., Papassotiropoulos, A., & Milnik, A. (2017). Exhaustive search for epistatic effects on the human methylome. *Scientific Reports*, *7*, 13669. https://doi.org/10.1038/s41598-017-13256-9

Engreitz, J. M., Daigle Jr., B. J., Marshall, J. J., & Altman, R. B. (2010). Independent component analysis: mining microarray data for fundamental human gene expression modules. *Journal of Biomedical Informatics*, *43*(6), 932–944. https://doi.org/10.1016/j.jbi.2010.07.001

Erhardt, E. B., Rachakonda, S., Bedrick, E. J., Allen, E. A., Adaly, T., & Calhoun, V. D. (2011). Comparison of multi-subject ICA methods for analysis of fMRI data. *Human Brain Mapping*, *32*, 2075–2095. https://doi.org/10.1002/hbm.21170

Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F., & Nyberg, L. (2015). Neurocognitive architecture of working memory. *Neuron*, *88*(1), 33–46. https://doi.org/10.1016/j.neuron.2015.09.020

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review*, *1*(2), 293–314. https://doi.org/10.1093/nsr/nwt032

Ferguson, A. R., Nielson, J. L., Cragin, M. H., Bandrowski, A. E., & Martone, M. E. (2014). Big data from small data: data-sharing in the “long tail” of neuroscience. *Nature*

Publishing Group, *17*(11), 1442–1447. https://doi.org/10.1038/nn.3838

Fox, P., & Hendler, J. (2011). Changing the equation on scientific data visualization. *Science*, *331*(6018), 705–709. https://doi.org/10.1126/science.1197962

Frane, A. V. (2016). False discovery rate control is not always a replacement for Bonferroni adjustment (Letter commenting on: J Clin Epidemiol. 2014;67:850-7). *Journal of Clinical Epidemiology*, *69*, 263. https://doi.org/10.1016/j.jclinepi.2015.03.025

Freytag, V., Carrillo-Roa, T., Milnik, A., Sämann, P. G., Vukojevic, V., Coynel, D., Demougin, P., Egli, T., Gschwind, L., Jessen, F., Loos, E., Maier, W., Riedel-Heller, S. G., Scherer, M., Vogler, C., Wagner, M., Binder, E. B., de Quervain, D. J.-F., & Papassotiropoulos, A. (2017). A peripheral epigenetic signature of immune system genes is linked to neocortical thickness and memory. *Nature Communications*, *8*, 15193. https://doi.org/10.1038/ncomms15193

Geib, B. R., Stanley, M. L., Wing, E. A., Laurienti, P. J., & Cabeza, R. (2017). Hippocampal contributions to the large-scale episodic memory network predict vivid visual memories. *Cerebral Cortex*, *37*(2), 680–693. https://doi.org/10.1093/cercor/bhv272

Gelernter, J. (2015). Complex trait genetics and population genetics in psychiatry: a review of the methods. In T. Canli (Ed.), *The Oxford Handbook of Molecular Psychology* (pp. 16–24). New York, NY: Oxford University Press.

Geraci, J., Wilansky, P., de Luca, V., Roy, A., Kennedy, J. L., & Strauss, J. (2017). Applying deep neural networks to unstructured text notes in electronic medical records for phenotyping youth depression. *Evidence-Based Mental Health*, *20*(3), 83–88.

Glickman, M. E., Rao, S. R., & Schultz, M. R. (2014). False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *Journal*

of Clinical Epidemiology, *67*(8), 850–857. https://doi.org/10.1016/j.jclinepi.2014.03.012

Godsey, B. (2017). *Think like a data scientist*. Shelter Island, NY: Manning.

Goldstein, H. (2011). *Multilevel Statistical Models* (4th ed.). Chichester: John Wiley & Sons.

Goodkind, M., Eickhoff, S. B., Oathes, D. J., Jiang, Y., Chang, A., Jones-Hagata, L. B., Ortega, B. N., Zaiko, Y. V, Roach, E. L., Korgaonkar, M. S., Grieve, S. M., Galatzer-Levy, I., Fox, P. T., & Etkin, A. (2015). Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry*, *5797*(4), 305–315. https://doi.org/10.1001/jamapsychiatry.2014.2206

Goodman, A., Pepe, A., Blocker, A. W., Borgman, C. L., Cranmer, K., Crosas, M., Di Stefano, R., Gil, Y., Groth, P., Hedstrom, M., Hogg, D. W., Kashyap, V., Mahabal, A., Siemiginowska, A., & Slavkovic, A. (2014). Ten simple rules for the care and feeding of scientific data. *PLoS Computational Biology*, *10*(4), 1–5. https://doi.org/10.1371/journal.pcbi.1003542

Gratten, J., Wray, N. R., Keller, M. C., & Visscher, P. M. (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nature Neuroscience*, *17*(6), 782–90. https://doi.org/10.1038/nn.3708

Gudbjartsson, D. F., Helgason, H., Gudjonsson, S. A., Zink, F., Oddson, A., Gylfason, A., Besenbacher, S., Magnusson, G., Halldorsson, B. V, Hjartarson, E., Sigurdsson, G. T., Stacey, S. N., Frigge, M. L., Holm, H., Saemundsdottir, J., Helgadottir, H. T., Johannsdottir, H., Sigfusson, G., Thorgeirsson, G., Sverrisson, J. T., Gretarsdottir, S., Walters, G. B., Rafnar, T., Thjodleifsson, B., Bjornsson, E. S., Olafsson, S., Thorarinsdottir, H., Steingrimsdottir, T., Gutmundsdottir, T. S., Theodors, A., Jonasson, J. G., Sigurdsson, A., Bjornsdottir, G., Jonsson, J. J., Thorarensen, O.,

Ludvigsson, P., Gudbjartsson, H., Eyjolfsson, G. I., Sigurdardottir, O., Olafsson, I., Arnar, D. O., Magnusson, O. T., Kong, A., Masson, G., Thorsteinsdottir, U., Helgason, A., Sulem, P., & Stefansson, K. (2015). Large-scale whole-genome sequencing of the Icelandic population. *Nature Genetics*, *47*(5), 435–444. https://doi.org/10.1038/ng.3247

Harlow, L. L., & Oswald, F. L. (2016). Big data in psychology: introduction to the special issue. *Psychological Methods*, *21*(4), 447–457. https://doi.org/10.1037/met0000120

Hart, E. M., Barmby, P., LeBauer, D., Michonneau, F., Mount, S., Mulrooney, P., Poisot, T., Woo, K. H., Zimmerman, N. B., & Hollister, J. W. (2016). Ten simple rules for digital data storage. *PLoS Computational Biology*, *12*(10), e1005097. https://doi.org/10.1371/journal.pcbi.1005097

Haw, R., Hermjakob, H., D'Eustachio, P., & Stein, L. (2011). Reactome pathway analysis to enrich biological discovery in proteomics data sets. *Proteomics*, *11*(18), 3598–613. https://doi.org/10.1002/pmic.201100066

Heck, A., Fastenrath, M., Ackermann, S., Auschra, B., Bickel, H., Coynel, D., Gschwind, L., Jessen, F., Kaduszkiewicz, H., Maier, W., Milnik, A., Pentzek, M., Riedel-Heller, S. G., Ripke, S., Spalek, K., Sullivan, P., Vogler, C., Wagner, M., Weyerer, S., Wolfsgruber, S., de Quervain, D. J.-F., & Papassotiropoulos, A. (2014). Converging genetic and functional brain imaging evidence links neuronal excitability to working memory, psychiatric disease, and brain activity. *Neuron*, *81*(5), 1203–1213. https://doi.org/10.1016/j.neuron.2014.01.010

Heck, A., Milnik, A., Vukojevic, V., Petrovska, J., Egli, T., Singer, J., Escobar, P., Sengstag, T., Coynel, D., Freytag, V., Fastenrath, M., Demougin, P., Loos, E., Hartmann, F., Schicktanz, N., Bizzini, B. D., Vogler, C., Kolassa, I.-T., Wilker, S., Elbert, T., Schwede,

T., Beisel, C., Beerenwinkel, N., de Quervain, D. J.-F., & Papassotiropoulos, A. (2017). Exome sequencing of healthy phenotypic extremes links TROVE2 to emotional memory and PTSD. *Nature Human Behavior*, *1*(4), 81. https://doi.org/10.1038/s41562-017-0081

Hemani, G., Theocharidis, A., Wei, W., & Haley, C. (2011). EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinformatics*, *27*(11), 1462–1465. https://doi.org/10.1093/bioinformatics/btr172

Holmes, A. P., & Friston, K. J. (1997). Generalisability, random effects & population inference. *Neuroimage*, *7*, 754.

Holzinger, A., Dehmer, M., & Jurisica, I. (2014). Knowledge discovery and interactive data mining in bioinformatics - state-of-the-art, future challenges and research directions. *BMC Bioinformatics*, *15*(Suppl 6), 1–9. https://doi.org/10.1186/1471-2105-15-S6-I1

Hyvärinen, A. (2013). Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1984), 1–19.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks*, *13*(4–5), 411–430. https://doi.org/10.1016/S0893-6080(00)00026-5

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Ioannidis, J. P. A., & Khoury, M. J. (2013). Are randomized trials obsolete or more important than ever in the genomic era? *Genome Medicine*, *5*(32), 1–3.

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). New York, NY: Springer.

Jones, D. K., Knösche, T. R., & Turner, R. (2013). White matter integrity, fiber count, and

other fallacies: the do's and don'ts of diffusion MRI. *NeuroImage*, *73*, 239–254. https://doi.org/10.1016/j.neuroimage.2012.06.081

Kane, M. J., Emerson, J. W., & Weston, S. (2013). Scalable strategies for computing with massive data. *Journal of Statistical Software*, *55*(14), 1–19. https://doi.org/10.18637/jss.v055.i14

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research*, *42*(Database issue), D199-205. https://doi.org/10.1093/nar/gkt1076

Kehrer, J., & Hauser, H. (2013). Visualization and visual analysis of multifaceted scientific data: a survey. *IEEE Transactions on Visualization and Computer Graphics*, *19*(3), 495–513. https://doi.org/10.1109/TVCG.2012.110

Kitts, A., Phan, L., Ward, M., & Holmes, J. B. (2014). The database of short genetic variation (dbSNP). In *The NCBI Handbook* (2nd ed.). Bethesda (MD): National Center for Biotechnology Information (US).

Kleppner, D., & Sharp, P. A. (2009). Research data in the digital age. *Science*, *325*(5939), 368. https://doi.org/10.1126/science.1178927

Klose, R. J., & Bird, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences*, *31*(2), 89–97. https://doi.org/10.1016/j.tibs.2005.12.008

Kong, W., Vanderburg, C. R., Gunshin, H., Rogers, J. T., & Huang, X. (2008). A review of independent component analysis application to microarray gene expression data. *Biotechniques*, *45*(5). https://doi.org/10.2144/000112950

Krapohl, E., Euesden, J., Zabaneh, D., Pingault, J.-B., Rimfeld, K., von Stumm, S., Dale, P. S., Breen, G., O'Reilly, P. F., & Plomin, R. (2016). Phenome-wide analysis of genome-wide polygenic scores. *Molecular Psychiatry*, *21*, 1188–1193.

https://doi.org/10.1038/mp.2015.126

Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). A primer on theory-driven web scraping: automatic extraction of big data from the internet for use in psychological research. *Psychological Methods*, *21*(4), 475–492. https://doi.org/10.1037/a0033269

Landhuis, E. (2017). Big brain, big data. *Nature*, *541*, 559–561. https://doi.org/10.1038/541559a

Lazar, N. (2016). The big picture: Imagining genetics: a tale of two modalities. *Chance*, *29*(2), 61–63. https://doi.org/10.1080/09332480.2016.1181968

Lessov-Schlaggar, C. N., Rubin, J. B., & Schlaggar, B. L. (2016). The fallacy of univariate solutions to complex systems problems. *Frontiers in Neuroscience*, *10*, 1–6. https://doi.org/10.3389/fnins.2016.00267

Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Reviews Genetics*, *3*, 662–673. https://doi.org/10.1038/nrg887

Loeffler, M., Engel, C., Ahnert, P., Alfermann, D., Arelin, K., Baber, R., Beutner, F., Binder, H., Brähler, E., Burkhardt, R., Ceglarek, U., Enzenbach, C., Fuchs, M., Glaesmer, H., Girlich, F., Hagendorff, A., Häntzsch, M., Hegerl, U., Henger, S., Hensch, T., Hinz, A., Holzendorf, V., Husser, D., Kersting, A., Kiel, A., Kirsten, T., Kratzsch, J., Krohn, K., Luck, T., Melzer, S., Netto, J., Nüchter, M., Raschpichler, M., Rauscher, F. G., Riedel-Heller, S. G., Sander, C., Scholz, M., Schönknecht, P., Schroeter, M. L., Simon, J.-C., Speer, R., Stäker, J., Stein, R., Stöbel-Richter, Y., Stumvoll, M., Tarnok, A., Teren, A., Teupser, D., Then, F. S., Tönjes, A., Treudler, R., Villringer, A., Weissgerber, A., Wiedemann, P., Zachariae, S., Wirkner, K., & Thiery, J. (2015). The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply

phenotyped adults in Germany. *BMC Public Health*, *15*, 691. https://doi.org/10.1186/s12889-015-1983-z

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*, 869–878. https://doi.org/10.1038/nature06976

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*, 150–157.

Luhmann, M. (2017). Using big data to study subjective well-being. *Current Opinion in Behavioral Sciences*, *18*, 28–33. https://doi.org/10.1016/j.cobeha.2017.07.006

Ma, Y., & Zhu, L. (2013). A review on dimension reduction. *International Statistical Review*, *81*(1), 1–18. https://doi.org/10.1111/j.1751-5823.2012.00182.x.A

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., Pendlington, Z. M., Welter, D., Burdett, T., Hindorff, L., Flicek, P., Cunningham, F., & Parkinson, H. (2017). The new NHGRI-EBI catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*, *45*(D1), D896–D901. https://doi.org/10.1093/nar/gkw1133

Matheson, S. (2017). Sorting out complex thoughts and messy emotions. *Cell*, *169*(7), 1157. https://doi.org/10.1016/j.cell.2017.06.004

McNab, F., Zeidman, P., Rutledge, R. B., Smittenaar, P., Brown, H. R., Adams, R. A., & Dolan, R. J. (2015). Age-related changes in working memory and the ability to ignore distraction. *Proceedings of the National Academy of Sciences*, *112*(20), 6515–6518. https://doi.org/10.1073/pnas.1504162112

Medland, S. E., Jahanshad, N., Neale, B. M., & Thompson, P. M. (2014). Whole-genome analyses of whole-brain data: working within an expanded search space. *Nature Neuroscience*, *17*(6), 791–800. https://doi.org/10.1038/nn.3718

Milnik, A., Heck, A., Vogler, C., Heinze, H.-J., de Quervain, D. J.-F., & Papassotiropoulos, A. (2012). Association of KIBRA with episodic and working memory: a meta-analysis. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *159B*(8), 958–969. https://doi.org/10.1002/ajmg.b.32101

Minzenberg, M. J., Laird, A. R., Thelen, S., Carter, C. S., & Glahn, D. C. (2010). Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia. *Archives of General Psychiatry*, *66*(8), 811–822. https://doi.org/10.1001/archgenpsychiatry.2009.91

Mladenić, D. (2006). Feature Selection for Dimensionality Reduction. In C. Saunders, M. Grobelnik, S. Gunn, & J. Shawe-Taylor (Eds.), *Subspace, Latent Structure and Feature Selection: Statistical and Optimization Perspectives Workshop, SLSFS 2005, Bohinj, Slovenia, February 23-25, 2005, Revised Selected Papers* (pp. 84–102). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/11752790_5

Monteith, S., Glenn, T., Geddes, J., & Bauer, M. (2015). Big data are coming to psychiatry: a general introduction. *International Journal of Bipolar Disorders*, *3*(21), 1–11. https://doi.org/10.1186/s40345-015-0038-9

Mooney, M. A., & Wilmot, B. (2015). Gene set analysis: a step-by-step guide. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *168*(7), 517–527. https://doi.org/10.1002/ajmg.b.32328

Munafò, M. R., & Flint, J. (2014). The genetic architecture of psychophysiological phenotypes. *Psychophysiology*, *51*, 1331–1332. https://doi.org/10.1111/psyp.12355

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E.-J., Ware, J. W., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behavior*, *1*, 21.

https://doi.org/10.1038/s41562-016-0021

Murell, P. (2006). *R Graphics*. (J. Lafferty, D. Madigan, F. Murtagh, & P. Smyth, Eds.). Boca Raton, FL: Chapman & Hall/CRC.

Murray, S. (2013). *Interactive data visualization for the web*. (M. Blanchette, Ed.). Sebastopol, CA: O'Reilly.

National Academy of Sciences. (2009). *On being a scientist: a guide to responsible conduct in research* (3rd ed.). Washington DC: National Academies Press (US). https://doi.org/10.17226/12192

Nelson, E. K., Piehler, B., Eckels, J., Rauch, A., Bellew, M., Hussey, P., Ramsay, S., Nathe, C., Lum, K., Krouse, K., Stearns, D., Connolly, B., Skillman, T., & Igra, M. (2011). LabKey server: an open source platform for scientific data integration, analysis and collaboration. *BMC Bioinformatics*, *12*(71), 1–23. https://doi.org/10.1186/1471-2105-12-71

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Sciences*, *7*(6), 615–631. https://doi.org/10.1177/1745691612459058

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Papassotiropoulos, A., & de Quervain, D. J.-F. (2011). Genetics of human episodic memory: dealing with complexity. *Trends in Cognitive Sciences*, *15*(9), 381–387. https://doi.org/10.1016/j.tics.2011.07.005

Papassotiropoulos, A., & de Quervain, D. J.-F. (2015). Failed drug discovery in psychiatry: time for human genome-guided solutions. *Trends in Cognitive Sciences*, *19*(4), 183–187. https://doi.org/10.1016/j.tics.2015.02.002

Papassotiropoulos, A., Stefanova, E., Vogler, C., Gschwind, L., Ackermann, S., Spalek, K.,

Rasch, B., Heck, A., Aerni, A., Hanser, E., Demougin, P., Huynh, K.-D., Luechinger, R., Klarhöfer, M., Novakovic, I., Kostic, V., Boesiger, P., Scheffler, K., & de Quervain, D. J.-F. (2013). A genome-wide survey and functional brain imaging study identify CTNNBL1 as a memory-related gene. *Molecular Psychiatry*, *18*(2), 255–263. https://doi.org/10.1038/mp.2011.148

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. P. (2014). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, *108*(6), 934–952. https://doi.org/10.1037/pspp0000020

Pers, T. H. (2016). Gene set analysis for interpreting genetic studies. *Human Molecular Genetics*, *25*(R2), R133–R140. https://doi.org/10.1093/hmg/ddw249

Pessoa, L. (2017). A network model of the emotional brain. *Trends in Cognitive Sciences*, *21*(5), 357–371. https://doi.org/10.1016/j.tics.2017.03.002

Poldrack, R. A. (2012). The future of fMRI in cognitive neuroscience. *NeuroImage*, *62*(2), 1216–1220. https://doi.org/10.1016/j.neuroimage.2011.08.007

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafo, M. R., Nichols, T. E., Poline, J.-B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*, 115–126. https://doi.org/doi:10.1038/nrn.2016.167

Poldrack, R. A., & Farah, M. J. (2015). Progress and challenges in probing the human brain. *Nature*, *526*, 371–379. https://doi.org/10.1038/nature15692

Poline, J.-B., Breeze, J. L., & Frouin, V. (2015). Imaging Genetics with fMRI. In K. Uludag, K. Ugurbil, & L. Berliner (Eds.), *fMRI: From Nuclear Spins to Brain Functions* (1st ed., pp. 699–738). New York, NY: Springer. https://doi.org/10.1007/978-1-4899-7591-1

R Core Team. (2013). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Ram, K. (2013). Git can facilitate greater reproducibility and increased transparency in science. *Source Code for Biology and Medicine*, *8*(7), 1–8. https://doi.org/10.1186/1751-0473-8-7

Schobel, J., Pryss, R., & Reichert, M. (2015). Using smart mobile devices for collecting structured data in clinical trials: results from a large-scale case study. In *IEEE 28th International Symposium on Computer-Based Medical Systems* (pp. 13–18). https://doi.org/10.1109/CBMS.2015.69

Schwartz, S. J., Lilienfeld, S. O., Meca, A., & Sauvigné, K. C. (2016). The role of neuroscience within psychology: a call for inclusiveness over exclusiveness. *American Psychologist*, *71*(1), 52–70. https://doi.org/10.1037/a0039678

Sejnowski, T. J., Churchland, P. S., & Movshon, J. A. (2014). Putting big data to good use in neuroscience. *Nature Neuroscience*, *17*(11), 1440–1441. https://doi.org/10.1038/nn.3839

Shen, J., Li, Z., Song, Z., Chen, J., & Shi, Y. (2017). Genome-wide two-locus interaction analysis identifies multiple epistatic SNP pairs that confer risk of prostate cancer: A cross-population study. *International Journal of Cancer*, *140*(9), 2075–2084. https://doi.org/10.1002/ijc.30622

Spiegelhalter, B. D. J. (2014). The future lies in uncertainty. *Science*, *345*(6194), 264–265. https://doi.org/10.1126/science.1251122

Stafford, T., & Dewar, M. (2014). Tracing the Trajectory of Skill Learning With a Very Large Sample of Online Game Players. *Psychological Science*, *25*(2), 511–518. https://doi.org/10.1177/0956797613511466

Stafford, T., & Haasnoot, E. (2017). Testing sleep consolidation in skill learning: a field

study using an online game. *Topics in Cognitive Science*, *9*, 485–496. https://doi.org/10.1111/tops.12232

Sullivan, P. F., & Posthuma, D. (2014). Biological pathways and networks implicated in psychiatric disorders. *Current Opinion in Behavioral Sciences*, *2*, 58–68. https://doi.org/10.1016/j.cobeha.2014.09.003

Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology*, *15*(3), e2000797. https://doi.org/10.1371/journal.pbio.2000797

The Gene Ontology Consortium. (2013). Gene Ontology annotations and resources. *Nucleic Acids Research*, *41*(D1), D530–D5535. https://doi.org/10.1093/nar/gks1050

Torous, J., Kiang, M. V., Lorme, J., & Onnela, J.-P. (2016). New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR Mental Health*, *3*(2), e16. https://doi.org/10.2196/mental.5165

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.

Tyner, C., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Eisenhart, C., Fischer, C. M., Gibson, D., Gonzalez, J. N., Guruvadoo, L., Haeussler, M., Heitner, S., Hinrichs, A. S., Karolchik, D., Lee, B. T., Lee, C. M., Nejad, P., Raney, B. J., Rosenbloom, K. R., Speir, M. L., Villarreal, C., Vivian, J., Zweig, A. S., Haussler, D., Kuhn, R. M., & Kent, W. J. (2017). The UCSC Genome Browser database: 2017 update. *Nucleic Acids Research*, *45*(D1), 626–634. https://doi.org/10.1093/nar/gkw1134

Unwin, A. (2008). Good graphics? In C. Chen, A. Unwin, & W. Härdle (Eds.), *Handbook of Data Visualization* (pp. 57–78). Berlin Heidelberg: Springer.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

van Horn, J. D., & Toga, A. W. (2014). Human neuroimaging as a "big data" science. *Brain Imaging and Behavior*, *8*(2), 323–331. https://doi.org/10.1007/s11682-013-9255-y

van Rossum, G. (1995). *Python tutorial, Technical Report CS-R9526*. Amsterdam, Netherlands.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Miklos, G. L. G., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., Mckusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R.-R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V, Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z. Y., Wang, A., Wang, X., Wang, J., Wei, M.-H., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S. C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L.,

Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.-H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.-H., Coyne, M., Dahlke, C., Mays, A. D., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., Mcdaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., & Zhu, X. (2001). The Sequence of the Human Genome. *Science*, *291*(5507), 1304–1351. https://doi.org/10.1126/science.1058040

Visscher, P. M., Brown, M. a, McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *American Journal of Human Genetics*, *90*(1), 7–24. https://doi.org/10.1016/j.ajhg.2011.11.029

Vogler, C., Gschwind, L., Coynel, D., Freytag, V., Milnik, A., Egli, T., Heck, A., de Quervain, D. J.-F., & Papassotiropoulos, A. (2014). Substantial SNP-based heritability estimates for working memory performance. *Translational Psychiatry*, *4*(9), e438. https://doi.org/10.1038/tp.2014.81

Wang, K., Li, M., & Hakonarson, H. (2010). Analysing biological pathways in genome-wide association studies. *Nature Reviews Genetics*, *11*(12), 843–854. https://doi.org/10.1038/nrg2884

Weaver, I. C. G., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Seckl, J. R., Dymov, S., Szyf, M., & Meaney, M. J. (2004). Epigenetic programming by maternal behavior. *Nature Neuroscience*, *7*(8), 847–854. https://doi.org/10.1038/nn1276

Wei, W.-H., Hemani, G., & Haley, C. S. (2014). Detecting epistasis in human complex traits. *Nature Reviews Genetics*, *15*(11), 722–733. https://doi.org/10.1038/nrg3747

Westfall, P. H., & Young, S. S. (1993). *Resampling-based multiple testing: examples and methods for p-value adjustment*. New York, NY: John Wiley.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer.

Wilson, G., Aruliah, D. A., Brown, C. T., Hong, N. P. C., Davis, M., Guy, R. T., Haddock, S. H. D., Huff, K. D., Mitchell, I. M., Plumbley, M. D., Waugh, B., White, E. P., & Wilson, P. (2014). Best practices for scientific computing. *PLoS Biology*, *12*(1), e1001745. https://doi.org/10.1371/journal.pbio.1001745

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. *PLoS Computational Biology*, *13*(6), e1005510. https://doi.org/10.1371/journal.pcbi.1005510

Woo, C.-W., Chang, L. J., Lindquist, M. A., & Wager, T. D. (2017). Building better biomarkers: brain models in translational neuroimaging. *Nature Neuroscience*,

20(3), 365–377. https://doi.org/10.1038/nn.4478

Wray, N. R. (2005). Allele frequencies and the r2 measure of linkage disequilibrium: impact on design and interpretation of association studies. *Twin Research and Human Genetics*, *8*(2), 87–94. https://doi.org/10.1375/twin.8.2.87

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*, 665–670. https://doi.org/10.1038/nmeth.1635

Yarkoni, T., Poldrack, R. A., Van Essen, D. C., & Wager, T. D. (2010). Cognitive neuroscience 2.0: building a cumulative science of human brain function. *Trends in Cognitive Sciences*, *14*(11), 489–496. https://doi.org/10.1016/j.tics.2010.08.004

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning (in press). *Perspectives in Psychological Science*.

Zhang, T., & Meaney, M. J. (2010). Epigenetics and the environmental regulation of the genome and its function. *Annual Review of Psychology*, *61*, 439–466. https://doi.org/10.1146/annurev.psych.60.110707.163625

7. Declaration by candidate

I declare herewith that I have independently carried out the PhD-thesis entitled "How complex analyses of large multidimensional datasets advance psychology – examples from large-scale studies on behavior, brain imaging, and genetics". This thesis consists of original research articles that have been written in cooperation with the enlisted co-authors and have been published in peer-reviewed scientific journals or are in preparation for publication / submitted for publication. Only allowed resources were used and all references used were cited accordingly.

Date: _____

Signature: _____