NeuroPredictome: A Data-Driven Predictome Linking Neuroimaging to Phenotype

Syed Fahad Sultan*, Lilianne Mujica-Parodi*†, Steven Skiena*

* Stony Brook University, Stony Brook, NY 11794, USA

† Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA 02129, USA

***Abstract*—Neuroimaging studies generally provide evidence only on a narrow aspects of the human brain function, suffering from small sample sizes and are hard to reproduce. These factors severely limit synthesis of neuroimaging findings and our ability to reach a global view of human brain organization, mapping and decoding. In this paper, we present a novel prediction based framework called *Neuropredictome* that allows identification of statistically significant linkages between phenotypes and neuroimaging features on UK-Biobank data. We evaluate phenotype linkage to brain fMRI activity on 4926 variables pertaining to the health, physiology, psychology, social and economic state for 19,831 subjects. We corroborate our identified regions of the brain with previous work by providing a novel quantitative evaluation of how well our results align with existing meta-analyses of 14,371 published neuroimaging research articles. Our analysis is presented as a public resource at https://neuropredictome.com providing an interpretable view of human brain organization and decoding, to assist in hypothesis generation and evaluating future studies.**

***Index Terms*—Neuroimaging, fMRI, natural language processing.**

I. Introduction

Each year, thousands of neuroimaging studies explore the links between brain, behavior, and (primarily) psychiatric and neurological disease. Individual studies however themselves, suffer from small sample sizes and hence seldom have the statistical power to establish fully trustworthy results [1], [2]. The median sample size of fMRI studies in 2015 was 28.5 subjects [3] and the 75th percentile of sample size in cognitive neuroscience journals published between 2011 to 2014 was 28 subjects [4]. In neuroimaging, sample sizes are small due to the financial cost of scans, which can exceed $1,200 USD per data point. But working with such small sample sizes yields low statistical power and an inflated false discovery rate. As a consequence, neuroimaging has been criticized for overestimating effect sizes [5] and concerns regarding reproducibility [6].

Meta-analyses and literature reviews try to resolve this problem by coalescing results from a wide number of studies. Coordinate-Based Meta-Analysis (CBMA) methods [7], [8], [9] assess the consistency of results across studies, comparing the observed spatial density of reported brain stereotactic coordinates to the null hypothesis of a uniform distribution. The latest automated CBMA methods such as Neurosynth [10] and Neuroquery [11] thus give excellent statistical power while enabling large scale analyses of brain-imaging studies. However, automated meta analyses frameworks come with their own set of limitations.

In this paper, we set out to address these concerns using a single large scale and public data set i.e. UK-Biobank to study and predict 4926 variables for 19,831 subjects. For each variable, we report our ability to decode it against naive baseline and the regions of the brain that are most relevant to that variable. Our results show highly significant results (at the 0.01 level) for 623 (12.6%) of Biobank variables, with another 1,046 (21.2%) variables significant at the 0.05 level, all after Bonferroni correction. Predictable variables include those associated with diet, substance use, fluid intelligence, and mental health. Our methodology fails to find relationships between measured fMRI activity and phenotypes such as bone fractures, ethnicity, and birth month, confirming the rigour of our statistical methods.

The model we construct for each BioBank phenotype identifies brain regions differentially active in positive subjects. These models provide a way to link phenotypes to the neuroimaging literature through the Neurosynth database, which collects brain map activation associated with over 14,371 peer-reviewed published studies. Over all Neurosynth papers and Uk-Biobank variables, the correlation between text similarity and fMRI similarity is strongest for highly significant categories (r=0.25), mildly strong for significant variables(r=0.18) and weak (r=0.09) for non-significant variables. This provides a quantitative evaluation of the brain activation maps identified by our classifier (in terms of learned weights) with existing activation maps and networks from published meta-analyses frameworks. We present these linkages between phenotype and the literature through fMRI activity at our website https://neuropredictome.com, where we anticipate it will prove a valuable resource to suggest connections between brain activity and disease.

In addition to providing a public, consistent, interpretable and reproducible global view of human brain organization and decoding, we hope our work to assist in hypothesis generation and providing data centric priors for grounding future studies.

The major contributions of this paper are as follows:

- *A novel prediction-based framework for linking phenotypes with fMRI data* – We have discovered statistically-rigorous linkages between over 600 cognitive, psychiatric, medical, and lifestyle factors and fMRI brain activity. Many of these linkages are novel, but further

978-1-6654-0126-5/21/$31.00 ©2021 IEEE

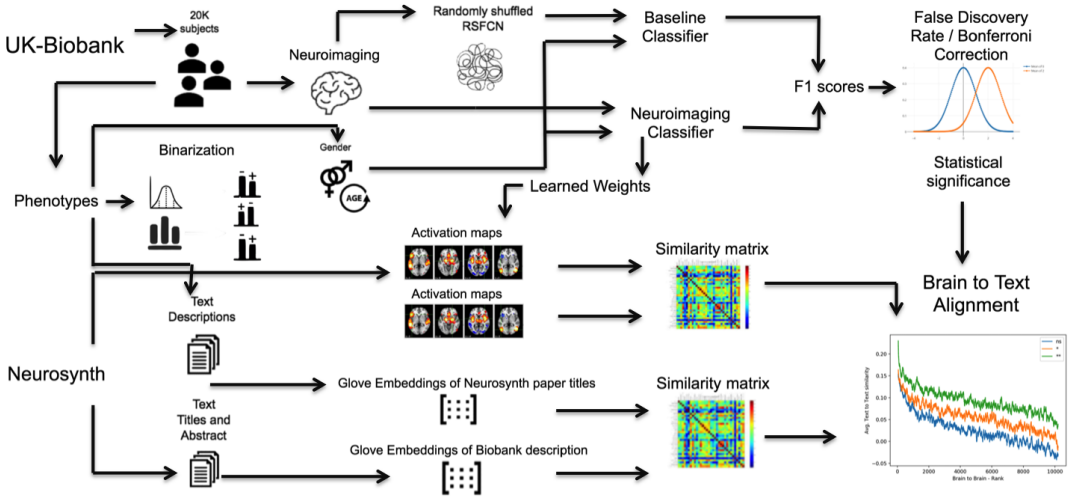

Fig. 1: **Schematic overview of NeuroPredictome**. For each neuroimaging modality, brain features of 19,831 subjects, along with age and sex, are used to train 4,926 classifiers, one for each binarized-phenotype, using 20-fold cross validation. Prediction scores from each classifier is compared to that of a baseline classifier to measure statistical significance. Weights learned by each classifier is used to generate an activation map for each phenotype that is then compared to activation maps reported in Neurosynth papers. UK-Biobank phenotypes and Neurosynth papers are similarly compared using text features and we then observe how well do similarities using text features align with similarities using brain features.

supported by literature-based evaluation.
The complete linkage data from this project is available at https://neuropredictome.com as a resource to motivate further research, providing strong baseline prediction scores for future work.

- *Evaluating the predictive power of different neuroimaging modalities* – The UK-Biobank contains subject data from four distinct neuroimaging modalities including structural (T1 and DTI) and functional (resting and task) imaging. There remains considerable debate in the field as to the relative power of these technologies for different tasks. We use the statistically-significant brain-phenotype linkages identified by Neuropredictome to establish that resting state functional connectivity (rfMRI) broadly yields the best prediction scores and predicts a larger number of phenotypes compared to structural DTI, T1-weighted or task fMRI.
- *Evaluating data representations for neuroimaging* – We perform a systematic evaluation of four different feature representations for fMRI data, employing the scale supported by our testbed enables use to make meaningful evaluations. We show that the relatively simple Resting State Functional Connectivity Networks (RSFCN) representation coupled with Logistic Regression (LR) as a classification model [12], [13] together exhibit predictive power comparable to more sophisticated deep learning representations and classifier – with substantially reduced computational costs.
- *Literature-based evaluation of neuroimaging-phenotypic linkage* – False positives through spurious correlations plague large-scale association studies such as ours. We have employed a novel literature-based evaluation to assess these linkages, which is of independent interest. We compare the brain maps implicit in our predictive model for each phenotype with external brain maps from the neuroscience literature. We demonstrate that the NLP text similarity between phenotype name and study title align surprisingly well with the statistical significance of the phenotype-fMRI linkage.

This paper is organized as follows. Section II presents the computational methodology behind Neuropredictome and details the UK Biobank dataset it is trained and evaluated on. Section III highlights the results of our analysis: both the statistical significance of important phenotypes and general observations on fMRI data representations and literature linkage. Our conclusions and proposals for future work are highlighted in Section IV.

II. METHODOLOGY

Our general strategy, as shown in Figure 1, starts with a large-scale hypothesis-neutral dataset to identify linkages between brain and phenotypes, linkages that are then cross-validated against meta-analyses from the neuroimaging literature. Each section to follow details individual component of our pipeline.

A. Data

The UK Biobank [14] is a prospective epidemiological study that recruited 500,000 adults between 2006-2010 [15]. Here, we considered an initial release of around 19,831 subjects with both structural MRI and functional MRI (both rest and task) data as well as Diffusion Tensor Imaging. Participants ages

range from 40 to 69 years of age at baseline recruitment. In addition to brain imaging data for these subjects, the study also collects data from extensive questionnaires, physical and cognitive measures, and biological samples (including genotyping). Following are details of the neuroimaging modalities we used for our experiments:

- *Resting State fMRI (rsfMRI)* – Both structural MRI and rs-fMRI were acquired on harmonized Siemens 3T Skyra scanners at four UK Biobank imaging centres (Cheadle, Manchester, Newcastle, and Reading). The structural MRI was 1.0mm isotropic. The rs-fMRI was 2.4mm isotropic with TR of 0.735s and 490 frames per run (6 min). Each subject had one rs-fMRI run.
- *Task functional Magnetic Resonance Imaging (tfMRI)* – Task functional MRI uses the same measurement technique as resting-state fMRI, while the subject performs a particular task or experiences a sensory stimulus. The task used in UKBiobank is the Hariri faces/shapes "emotion" task [16], [17]. The participants are presented with three faces or three shapes in two rows and they had to match the stimulus in the rows. The features from task fMRI scans used for classification were also functional connectivity weights, comparable to those computed for resting state fMRI.
- *Diffusion Tensor Imaging (DTI)* – DTI measures the ability of water molecules to move within their local tissue environment. Water diffusion is measured along a range of orientations, providing two types of useful information: i) Local voxel-wise estimates ii) Long-range estimates based on tract-tracing. As features to our classifier, in our experiments, we use for each tract, and for each output image type, the weighted-mean value of the probabilistic tractography output.
- *Structural T1-weighted imaging* – This structural technique gives a high-resolution depiction of brain anatomy, having strong contrast between grey and white matter, reflecting differences in the interaction of water with surrounding tissue. This modality provides derived fields primarily relating to volumes of brain tissues and structures. It is also critical for calculations of cross-subject and cross-modality alignments, needed in order to process all other brain modalities. The data available to us was at the resolution of 1 mm isotropic.

B. Preprocessing Phenotypes

In this study, we restricted ourselves to the problem of binary classification. Therefore, we binarize each of 5,034 phenotypes. To aid in interpretability, the positive class (disease group) is always the smaller of the two classes and the negative class (control group) is always the majority class. Without binarization, for certain phenotypes we would have to carry out regression and for others classification. This would have made the results harder to interpret. This binarization step ensures consistency in reporting across prediction of all phenotypes, nominal or ordinal.

Each nominal phenotype was converted into a variant of type one vs. rest for each value it takes on. Additionally, since not all phenotypes were available for all the subjects, an *available vs. not-available* variant was also created. Each of the ordinal phenotypes were binarized into 3 variants, one with respect to the three quartiles: i) 25th quartile vs. rest ii) 50th quartile vs. rest and iii) 75th quartile vs. rest. Once the binarized variants of all the phenotypes were computed, variants that were available for at least 1000 subjects and for which the minority class was at least one percent of all the subjects were retained and the remaining were dropped. Following these procedures, we finally ended up with 4926 phenotype variants.

C. Data Representations

We evaluate the performance of four distinct fMRI data representations for predictive models:

- *RSFCN* – Resting State Functional Connectivity Network (RSFCN) measures the congruity between different regions of the brain while the participants are lying at rest without any explicit task. Each entry of the RSFCN matrix corresponds to the strength of the functional connectivity between two brain regions, computed using the full Pearson correlation coefficient of the corresponding pair of time series. RSFC data of each participant was summarized as an $N \quad N$ matrix, where N is the number of brain ROIs.The entries of the RSFCN matrix were used as features to predict behavioral and demographic measures in individual participants. For our experiments, We use 55 ROIs obtained using spatial-ICA [18], [15], [19] and removing artefacts [15].
- *FNN* – Fully-connected neural networks (FNNs) belong to a generic class of feedforward neural networks [20]. An FNN takes in vector data as an input and outputs a vector. An FNN consists of several fully connected layers. In this work, the inputs to the FNN are the vectorized RSFCN (i.e., lower triangular entries of the RSFC matrices) and the outputs are the behavioral or demographic variables we seek to predict.
- *BrainNetCNN* – BrainNetCNN [21] is a specially designed DNN for connectivity data. BrainNetCNN allows the application of convolution to connectivity data, resulting in significantly less trainable parameters than the FNN. In this work, the input to the BrainNetCNN is the RSFC matrix and the outputs are the behavioral or demographic variables we seek to predict.
- *GCNN* – Standard convolution applies to data that lies on a Euclidean grid. Graph convolution exploits the graph Laplacian in order to generalize the concept of standard convolution to data lying on nodes connected together into a graph. This allows the extension of the standard CNN to graph convolutional neural networks. Here we considered the innovative GCNN developed by Kipf and Welling [22] and extended to neuroimaging data by Parisot and colleagues [23].

D. Identifying Statistically Significant Predictions

1) Confound modeling: Confounds can be significant in addressing problems of unexplained variance and spurious correlations. Alfaro et al. [24] describe a set of possible confounds in UK-Biobank and show that imaging can be influenced by blood pressure, bone density, height and weight. Age and sex in particular are confirmed as one of the most important confounds that mediate large amounts of between-subject variance. Preliminary results of our experiments, in agreement with existing literature [24], [25], [12], [13], showed that fMRI scans can reveal the subject's age and sex with a high degree of accuracy. We corrected for age and sex by providing them as features to both our RSFCN classifier and the Baseline classifier. A detailed discussion on studies for which these confounds may be useful can be found in Barnes et al. [25].

2) Baseline Model: As baseline model, we used a Logistic Regression classifier with features age and sex. In order to have parity in terms of the number of features in our model and that of the baseline, we provided the baseline model with randomly shuffled fMRI features. In other words, the fMRI features available to the baseline model did not have the original subject to fMRI features thus encoding useless information yet preserving the overall properties of the fMRI features.

3) 20-fold Cross Validation: We create 20 different splits for our subjects into training and testing sets. The train:test split was 75:25. For each split, we trained age-sex matched classifiers on the training subjects and computed f1 scores from prediction of our RSFCN classifier and Naive Baseline model. This procedure gave us two distributions, one from our model and the other from baseline, of f1-scores, each constituent of 20 data points, one from each of the 20 folds. F-1 scores were used due to their robustness in case of highly imbalanced cases. For such skewed data, majority classifiers, which always pick the largest class, can achieve trivially high accuracy.

4) Computing Statistical significance: To measure how distinctly better our model did compared to a random baseline, we performed student t-tests on the distribution of F1-scores from our model and the mean score of the baseline. This yielded a p-value for the null hypothesis that the baseline model's mean prediction was sampled from the distribution of prediction scores of our model. We use this p-value and its associated t-test statistic as the two primary measures of quality of our models. Given the large number of phenotypes, we used Bonferroni correction to control for false discovery rates and to evaluate the quality or reliability of the accuracy score. The p-value scores in the results presented were after applying the Bonferroni correction.

E. Evaluation against Neurosynth

Neurosynth [10] is repository of activation maps generated by an automated meta-analysis. It associates activation maps with 14,371 published neuroimaging papers and also summarizes information from these activation maps. These summaries are with respect to coordinates in the standard MNI space and the general common terms found in the title and abstract of these papers. Neurosynth represents the state-of-the-art in the neuroimaging community's understanding of relationship between regions in the brain and a wide array of behavior and phenotypes. Here, we align knowledge gained from UK-Biobank with that in Neurosynth.

1) Neurosynth Brain to Biobank Brain mapping: Each of our 4,926 classifiers, trained against a particular phenotype or its variant, learns weights on the given features. These weights are used to compute an activation map. This is carried out in the following three steps. First, for each classifier, we average learned weights from each fold of the 20-fold cross validation. Second, we map a learned weight for each edge in the connectivity graph matrix to the corresponding pairs of ROIs. Finally, the ROI weights are mapped onto voxels. Since ROIs are essentially sets of voxels, we assign to each voxel the weight of its respective ROI. In case of voxels that fall into more than one ROI, we assign it the average weight of all its corresponding ROIs. Once we have an activation map for each of our 4926 phenotype variants and an activation map for each of the 14,371 papers in Neurosynth, we compute a pairwise similarity matrix (4926,14371) between UK-Biobank phenotypes and Neurosynth paper embeddings, using Pearson correlation coefficient as the similarity metric.

2) Neurosynth text to Biobank text mapping: We compared text descriptions of the 4,926 UK-Biobank phenotypes with the text from abstract and title of the 14,371 papers in Neurosynth. We do so by treating the phenotype/variable descriptions in the official documentation as text documents. Similarly, we concatenate the title and abstract of the paper titles to get text documents for each of the Neurosynth papers. We then use a deep learning-based text embedding technique called Glove embeddings [26] to convert these documents from Neurosynth and UK-Biobank to 300-dimensional vectors. This gives us two matrices: 1) $4926 \quad 300$ dimensional matrix for UK-Biobank phenotypes and 2) $14371 \quad 300$ matrix for Neurosynth papers. We then compute pairwise distances for each pair of phenotype and Neurosynth paper using cosine distance to obtain a $4926 \quad 14371$ dimensional text similarity matrix.

3) Brain and text alignment: Once the two similarity matrices are computed, each giving an all-to-all comparison between Neurosynth papers and Biobank variables, one in text space and the other in brain space (cell (i, j) of each matrix representing the same phenotype i and Neurosynth paper j pair) we compute an alignment score between the two matrices. The alignment score was computed using Pearson correlation coefficient between the two matrices. This alignment score represents the degree of agreement between the two spaces i.e. if a paper-phenotype pair have highly similar activation maps, then that pair also has highly similar text descriptions.

III. RESULTS

A. Choosing a representation and a classifier

For classification, we needed a representation as features and a classifier. We evaluated Resting State Functional Con-

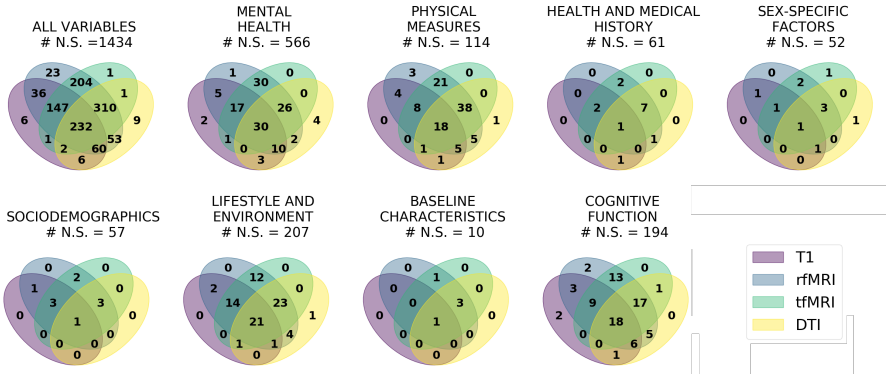

Fig. 2: **Neuropredictome identifies new brain-phenotype linkages, along with some expected ones**. Physical and lifestyle phenotypes, typically not looked at in light of neuroimaging, showed strong brain effects. Across neuroimaging modalities, resting state functional connectivity (rfMRI) broadly yields the best prediction scores and predicts a larger number of phenotypes compared to structural DTI, T1-weighted or task fMRI.

nectivity Networks (RSFCN) and Logistic Regression (LR) against other more representations and classifiers against more sophisticated deep learning representations and more involved. Our results reaffirmed findings in previous work [12], [13] that show RSFCN and LR to exhibit the most widely better predictive power. Results of our evaluation for picking the representation is given in Table I and our results for comparison between different classifiers is given in II.

B. Neuropredictome identifies brain-phenotype linkages, both expected and new within neuroimaging.

A summary of UK Biobank neuro-phenotype linkages is provided in Figure 2; representative variables of interest and their corresponding prediction statistics are shown in Table III. For all variables measuring cognitive performance, subjects who scored above the median were distinguished from subjects who did worse. Many fine-grained aspects of mental health, such as depression, anxiety, mood swings, manic episodes and poor appetite were also predicted from fMRI brain activity.

			Representation	RSFCN	FNN	BrainNetCNN	GCNN
	Variable	**Binarization**	**Metric**				
1	Sex		Accuracy	**0.916**	0.89	0.9	0.86
2	Age		Correlation	**0.599**	**0.599**	0.598	0.593
			MAE	4.826	4.899	**4.824**	4.895
3	Pairs Matching		Correlation	0.061	0.045	**0.067**	0.008
4	Fluid Intelligence		Correlation	**0.239**	**0.239**	0.235	0.232
5	Assessment Center		Accuracy	0.67	0.65	0.62	0.66
6	BMI	<75percentile	Accuracy	0.76	0.77	**0.78**	0.71
7	Fluid Inteligence	<50percentile	Accuracy	0.64	0.59	**0.66**	0.64
8	HeadMotion	<25 percentile	Accuracy	**0.8**	0.74	0.79	0.73
9	Volume of Grey Matter	<25percentile	Accuracy	**0.82**	0.78	0.75	0.79
10	Pulse Rate	<25perentile	Accuracy	**0.65**	0.63	0.6	0.61
11	Alchohol Intake	>=Weekly	Accuracy	**0.72**	0.71	0.68	0.69
12	Depression		Accuracy	0.69	0.67	**0.7**	0.65
13	Physical Activity	<=moderate	Accuracy	**0.63**	0.58	0.6	0.55
14	Ever Smoked		Accuracy	**0.6**	0.55	0.58	0.55

TABLE I: **Resting State Functional Connectivity Networks (RSFCN) yielded the best predictions scores for the selected set of representative phenotypes.** Deep learning based Fully Connected Neural Network (FNN), Brain Net Convolutional Neural Network (BrainNetCNN) and Graph Convolutional Neural Networks (GCNN) representations were used as baselines for comparison.

		Significant %		AUC ROC Significant		% Brain MRI significant		AUC ROC Sig. Brain MRI	
		LR	**SVM**	**LR**	**SVM**	**LR**	**SVM**	**LR**	**SVM**
1	RSFCN(16)	0.56	0.53	0.59	0.55	0.68	0.62	0.7	0.56
2	RSFCN(32)	0.55	0.55	0.58	0.54	0.68	0.61	0.68	0.62
3	RSFCN(64)	0.56	0.53	0.58	0.54	0.7	0.64	0.7	0.65
4	RSFCN(128)	0.56	0.54	0.56	0.55	0.69	0.6	0.72	0.61

TABLE II: **Logistic Regression (LR) identified a larger number of statistically significant phenotype-neuroimaging linkages compared to Support Vector Machines (SVM)**.

That psychiatric and cognitive phenotypes were linked to brain features was reassuring, but not surprising. However, one of our most striking results was that the physical and lifestyle phenotypes of greatest relevance to general medicine, while not typically assessed by neuroimaging studies, also showed brain effects. These included physical measures such as diet, Body Mass Index (BMI), and cardiovascular disorders such as angina, heart attack, strokes and blood pressure. It also included variables believed to be determinants of general well-being, such as quality of sleep, lack of social support system, experience of past trauma, satisfaction with family relationships, financial security, as well as leisure and social activities. Smoking, alcohol consumption and cannabis usage, even occasional, or in the past, was also found to alter brain activity enough for it to be strongly identifiable from resting state fMRI scans. The NP classifier was able to distinguish subjects who had never smoked cigarettes or cannabis from those who have tried it at some point, including participants who had already quit, suggesting a surprisingly long lasting footprint of addictive substances on the brain that endures long past consumption. For those that still actively consumed these substances, NP could predict the amount of consumption up to a coarse approximation. Importantly, the NP classifier passed critical "sanity checks" by failing to find relationships between brain measures and phenotypes such as bone fractures and month of birth. Likewise, brain measures did not predict ethnicity.

Comparing across neuroimaging modalities, our results

TABLE III: **Neuropredictome identifies brain-phenotype linkages, across different phenotype categories, that have gone ignored in neuroimaging along with some linkages that were to be expected.**. A linkage is considered strongly statistically significant if p-value <0.001 (**) and significant if p-value <0.01 (*) based on rejection of the null-hypothesis of association between F1-score distributions, from different folds of cross-validation, using Resting-State Functional Connectivity Network (RSFCN) and F1-score distribution from baseline classifier (Base) after Bonferroni correction.

Category	Variable	+ve pct	F1 (Base)	F1 (RFCN)	p-value	sig*
Health and medical history	Fractured/broken bones in last 5 years	0.083	0	0.00417	0.33	
Health and medical history	Fracture resulting from simple fall rest (+ve) vs Yes	0.429	0.496	0.501	0.405	
Health and medical history	Fractured bone site(s) Ankle (+ve) vs rest	0.0938	0.0178	0.0538	0.000241	
Health and medical history	Fractured bone site(s) Wrist (+ve) vs rest	0.181	0.0528	0.0678	0.0167	
Lifestyle and environment	Major dietary changes in the last 5 years rest (+ve) vs No	0.36	0.03	0.09	4.13e-18	**
Lifestyle and environment	Beef intake $>$ Less than once a week	0.394	0.0951	0.151	7.87e-16	**
Lifestyle and environment	Coffee intake $>$ 75th percentile	0.19	0	0.000832	0.00211	
Health and medical history	Diabetes diagnosed by doctor Yes (+ve) vs rest	0.0499	0	0.000375	0.33	
Lifestyle and environment	Number of cigarettes currently smoked daily $<$ 50th percentile	0.469	0.469	0.523	1.05e-06	**
Lifestyle and environment	Smoking status Previous (+ve) vs rest	0.336	0.102	0.132	1.37e-10	**
Lifestyle and environment	Ever taken cannabis Yes (+ve) vs No	0.215	0.0796	0.136	2.35e-12	**
Lifestyle and environment	Frequency of drinking alcohol $\leq$ Monthly or less	0.197	0.00114	0.0186	2.22e-09	**
Lifestyle and environment	Frequency of drinking alcohol $>$ 2 to 3 times a week	0.303	0.0701	0.123	8.97e-14	**
Cognitive Function	Fluid intelligence score $<$ 50th percentile	0.455	0.345	0.485	1.55e-23	**
Cognitive Function	Matrix Pattern Completion $<$ 50th pctl.	0.372	0.279	0.378	2.33e-17	**
Cognitive Function	Numeric Memory $<$ 50th percentile	0.338	0.0987	0.194	5.35e-17	**
Cognitive Function	Reaction Time $<$ 50th pctl.	0.492	0.603	0.614	2.39e-06	*
Cognitive Function	Trail making $<$ 50th percentile	0.493	0.611	0.626	8.52e-07	**
Cognitive Function	Tower rearranging $<$ 50th percentile	0.405	0.358	0.411	6.18e-11	**
Mental Health	Seen doctor (GP) for nerves, anxiety, tension or depression	0.321	0.104	0.138	2.05e-11	**
Mental Health	Bipolar and major depression status $>$ No Bipolar or Depression	0.291	0.114	0.18	7.23e-11	**
Mental Health	Manic/hyper symptoms None of the above (+ve) vs rest	0.43	0.391	0.429	9.18e-10	**
Mental Health	Family relationship satisfaction $>$ Very happy	0.322	0.0177	0.0415	9.69e-11	**
Mental Health	Financial situation satisfaction $>$ Very happy	0.408	0.301	0.325	4.83e-10	**
Mental Health	Leisure/social activities None of the above (+ve) vs rest	0.245	8e-05	0.01	2.9e-11	**
Mental Health	Victim of sexual assault $>$ Never	0.152	0.000532	0.00596	5.78e-06	*
Mental Health	Physically abused by family as a child $>$ Never true	0.199	0.000269	0.00468	5.61e-07	**
Physical measures	Body mass index (BMI) $>$ 75th percentile	0.248	0.000736	0.145	1.5e-23	**
Physical measures	Diastolic blood pressure, manual reading $>$ 75th percentile	0.229	0.0925	0.141	1.38e-07	**
Physical measures	IPAQ activity group $>$ moderate	0.397	0.103	0.172	6.12e-17	**
Lifestyle and environment	Sleeplessness / insomnia $\leq$ Never rarely	0.221	0.00345	0.011	2.5e-07	**
Lifestyle and environment	Daytime dozing / sleeping (narcolepsy) Sometimes (+ve) vs rest	0.205	0.00168	0.0176	7.54e-10	**
Lifestyle and environment	Sleep duration $<$ 25th percentile	0.244	0.000247	0.012	8.74e-10	**
Sociodemographics	Average total household income before tax $\leq$ 18,000	0.213	0.0138	0.0561	9.99e-15	**
Sociodemographics	Average total household income before tax $>$ 31,000	0.254	0.256	0.309	2.89e-13	**
Sociodemographics	Ethnic background African (+ve) vs rest	0.00199	0	0.0161	0.109	
Sociodemographics	Ethnic background Chinese (+ve) vs rest	0.00277	0	0.031	0.0153	
Sociodemographics	Ethnic background Indian (+ve) vs rest	0.00654	0	0.0295	0.00475	
Sociodemographics	Ethnic background rest (+ve) vs British	0.0744	0	0.0043	0.00122	
	UK Biobank assessment center Newcastle (+ve) vs Cheadle	0.156	0	0.246	1.2e-24	**

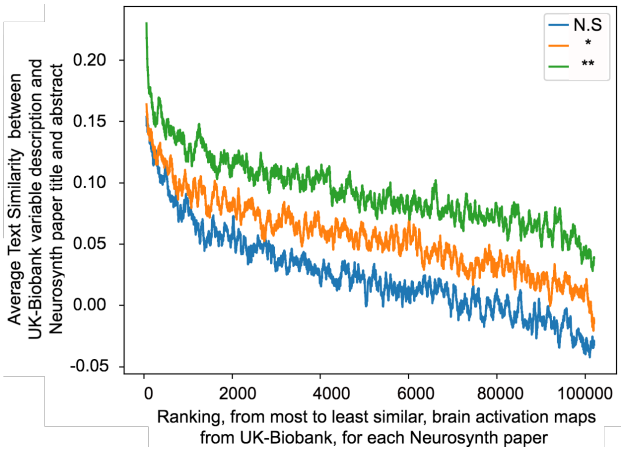

Fig. 3: **Neuropredictome results strongly align with results reported in 14,371 Neurosynth papers.** When Neuropredictome maps a Neurosynth paper to a UK-Biobank phenotype based on similarity in brain activity, then the phenotype-paper pair are also similar in terms of textual descriptions. The more confident Neuropredictome is about its predictions, the stronger the alignment. Most significant phenotypes (**) $r = 0.25$; moderately significant $r = 0.18$ (*); for non-significant (N.S) phenotypes $r = 0.09$.

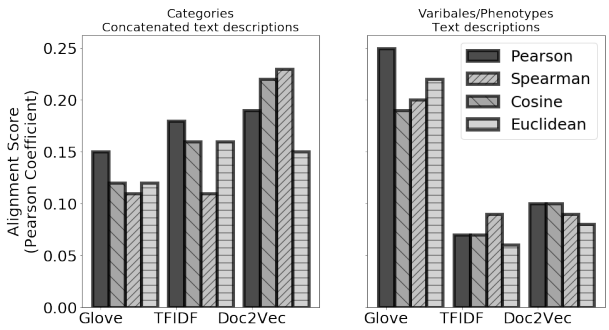

Fig. 4: **Glove text embeddings [26] over individual phenotype/variable descriptions and pearson as the brain similarity measure, together, yielded the best alignment score**. We conducted an extensive search over the hyperparameter search space to optimize for the best alignment scores. Other fixed hyperparameters were as follows: the brain space was MNI, the text-to-text similarity measure was cosine and the measure for the final alignment score was pearson correlation coefficient.

show Resting-state fMRI to broadly yield the best prediction scores and to predict the largest number of phenotypes as compared to structural DTI, T1-weighted or task fMRI. In line with previous findings [27], our results show that resting functional connectivity lends *additive* prediction power to structural neuroimaging and is a more generally informative fingerprint of the subject as compared to task networks [28].

Also of note for multi-site studies was the profound effect of imaging site. This was despite the fact that in order to maximize data compatibility across the different imaging centers, identical scanners were used with no major software or hardware updates throughout the study, with identical acquisition parameters, as well as identical post-processing pipelines. These results are in line with recent work by our group showing the influence of scanner-specific artifacts on neuroimaging data [29].

C. Brain maps learned by our classifier align with results from 14,371 published papers.

Figure 3 shows that Neuropredictome's classifiers from Biobank linkages were strongly supported by Neurosynth meta-analyses of the neuroimaging literature. It demonstrates

TABLE IV: **Activation Maps learned by Neuropredictome for a wide variety of phenotypes strongly align with relevant Neurosynth papers.** A curated list is given of UK-Biobank phenotypes and their corresponding top ranked Neurosynth paper with respect to similarity in reported brain activations and those identified by Neuropredictome. Corresponding text similarity between description of phenotype and title and abstract of top-ranked paper is also given.

ID	Biobank variable	pval	Text similarity	Neurosynth paper title	PubMed ID
20116	Smoking status > Never	2.15e-16	0.21	Nicotine withdrawal modulates frontal brain function during an affective Stroop task	21989805
20414	Frequency of drinking alcohol > 2 to 3 times a week	8.97e-14	0.2	Influence of cue exposure on inhibitory control and brain activation in patients with alcohol dependence.	22557953
5699	Fluid intelligence: arithmetic sequence recognition	5.52E-17	0.13	The interaction of process and domain in prefrontal cortex during inductive reasoning	25498406
20240	Numeric Memory - Maximum digits remembered correctly < 50th percentile	5.35e-17	0.17	Cross-modal processing in auditory and visual working memory	16154767
6772	Trail making - Interval between previous point and current one in numeric path >75 percentile	4.26E-14	0.22	The neural basis for establishing a focal point in pure coordination games.	22009019
2090	Seen doctor (GP) for nerves, anxiety, tension or depression	4.26E-14	0.36	Regional homogeneity associated with overgeneral autobiographical memory of first-episode treatment-naive patients with major depressive disorder in the orbitofrontal cortex	27923192
20425	Anxiety - Ever worried more than most people would in similar situation	2.01E-12	0.22	Metabolic and functional connectivity changes in mal de debarquement syndrome	23209584
20126	Bipolar II Disorder (+ve) vs rest	2.87E-11	0.36	Possible structural abnormality of the brainstem in unipolar depressive illness	16227541
21002	Weight > 75% percentile	6.97E-24	0.24	Widespread reward-system activation in obese women in response to pictures of high-calorie foods.	18413289
1220	Daytime dozing / sleeping (narcolepsy)	3.64E-13	0.13	Hippocampal activity mediates the relationship between circadian activity rhythms and memory in older adults.	26205911

that when our brain maps align well with Neurosynth brain maps (the low ranking categories to the left), the text descriptions of Biobank variables match well with paper titles in the literature. Further, the text similarity degrades as we move from variables ranked highly significant to significant to non-significant by NeuroPredictome, as should be the case if we make accurate linkages between fMRI images and categorical variables.

IV. Conclusion

We have presented a novel machine learning based predictive framework called NeuroPredictome (NP) that identifies new relationships between brain-based features and phenotypes. All identified associations between phenotypes and the brain can be accessed at https://neuropredictome.com. Future work should focus on studying better representations for fMRI analysis, and the question of whether non-linear models (including neural networks) can identify more statistically significant relations given the data and evaluation environments.

Acknowledgment

The research described in this paper was funded by the W. M. Keck Foundation (L.R.M.-P.), the White House Brain Research Through Advancing Innovative Technologies (BRAIN) Initiative (Grants NSFECCS1533257 and NSFNCS-FR 1926781 to L.R.M.-P.), and the US National Academies (Grant NAKFICB8 to L.R.M.-P.). This work was also partially supported by NSF grants IIS-1926781, IIS-1927227, IIS-1546113 and OAC-191952.

References

[1] K. S. Button, J. P. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. Robinson, and M. R. Munafò, "Power failure: why small sample size undermines the reliability of neuroscience," *Nature Reviews Neuroscience*, vol. 14, no. 5, pp. 365–376, 2013.

[2] T. D. Wager, M. Lindquist, and L. Kaplan, "Meta-analysis of functional neuroimaging data: current and future directions," *Social cognitive and affective neuroscience*, vol. 2, no. 2, pp. 150–158, 2007.

[3] R. A. Poldrack, C. I. Baker, J. Durnez, K. J. Gorgolewski, P. M. Matthews, M. R. Munafò, T. E. Nichols, J.-B. Poline, E. Vul, and T. Yarkoni, "Scanning the horizon: towards transparent and reproducible neuroimaging research," *Nature reviews neuroscience*, vol. 18, no. 2, p. 115, 2017.

[4] B. O. Turner, E. J. Paul, M. B. Miller, and A. K. Barbey, "Small sample sizes reduce the replicability of task-based fmri studies," *Communications Biology*, vol. 1, no. 1, pp. 1–10, 2018.

[5] H. R. Cremers, T. D. Wager, and T. Yarkoni, "The relation between statistical power and inference in fmri," *PloS one*, vol. 12, no. 11, p. e0184923, 2017.

[6] K. J. Gorgolewski and R. A. Poldrack, "A practical guide for improving transparency and reproducibility in neuroimaging research," *PLoS Biology*, vol. 14, no. 7, p. e1002506, 2016.

[7] T. Yarkoni, R. A. Poldrack, D. C. Van Essen, and T. D. Wager, "Cognitive neuroscience 2.0: building a cumulative science of human brain function," *Trends in cognitive sciences*, vol. 14, no. 11, pp. 489–496, 2010.

[8] P. T. Fox, L. M. Parsons, and J. L. Lancaster, "Beyond the single study: function/location metanalysis in cognitive neuroimaging," *Current opinion in neurobiology*, vol. 8, no. 2, pp. 178–187, 1998.

[9] J. D. Van Horn, S. T. Grafton, D. Rockmore, and M. S. Gazzaniga, "Sharing neuroimaging studies of human cognition," *Nature neuroscience*, vol. 7, no. 5, pp. 473–481, 2004.

[10] T. Yarkoni, R. A. Poldrack, T. E. Nichols, D. C. Van Essen, and T. D. Wager, "Large-scale automated synthesis of human functional neuroimaging data," *Nature methods*, vol. 8, no. 8, p. 665, 2011.

[11] J. Dockès, R. A. Poldrack, R. Primet, H. Gözükan, T. Yarkoni, F. Suchanek, B. Thirion, and G. Varoquaux, "Neuroquery, comprehensive meta-analysis of human brain mapping," *Elife*, vol. 9, p. e53385, 2020.

[12] T. He, R. Kong, A. J. Holmes, M. Nguyen, M. R. Sabuncu, S. B. Eickhoff, D. Bzdok, J. Feng, and B. T. Yeo, "Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics," *NeuroImage*, vol. 206, p. 116276, 2020.

[13] K. Dadi, M. Rahim, A. Abraham, D. Chyzhyk, M. Milham, B. Thirion, G. Varoquaux, A. D. N. Initiative *et al.*, "Benchmarking functional connectome-based predictive models for resting-state fmri," *NeuroImage*, vol. 192, pp. 115–134, 2019.

[14] C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray *et al.*, "Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age," *PLoS medicine*, vol. 12, no. 3, 2015.

[15] K. L. Miller, F. Alfaro-Almagro, N. K. Bangerter, D. L. Thomas, E. Yacoub, J. Xu, A. J. Bartsch, S. Jbabdi, S. N. Sotiropoulos, J. L. Andersson *et al.*, "Multimodal population brain imaging in the uk biobank prospective epidemiological study," *Nature neuroscience*, vol. 19, no. 11, p. 1523, 2016.

[16] A. R. Hariri, A. Tessitore, V. S. Mattay, F. Fera, and D. R. Weinberger, "The amygdala response to emotional stimuli: a comparison of faces and scenes," *Neuroimage*, vol. 17, no. 1, pp. 317–323, 2002.

[17] D. M. Barch, G. C. Burgess, M. P. Harms, S. E. Petersen, B. L. Schlaggar, M. Corbetta, M. F. Glasser, S. Curtiss, S. Dixit, C. Feldt *et al.*, "Function in the human connectome: task-fmri and individual differences in behavior," *Neuroimage*, vol. 80, pp. 169–189, 2013.

[18] C. F. Beckmann and S. M. Smith, "Probabilistic independent component analysis for functional magnetic resonance imaging," *IEEE transactions on medical imaging*, vol. 23, no. 2, pp. 137–152, 2004.

[19] F. Alfaro-Almagro, M. Jenkinson, N. K. Bangerter, J. L. Andersson, L. Griffanti, G. Douaud, S. N. Sotiropoulos, S. Jbabdi, M. Hernandez-Fernandez, E. Vallee *et al.*, "Image processing and quality control for the first 10,000 brain imaging datasets from uk biobank," *Neuroimage*, vol. 166, pp. 400–424, 2018.

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[21] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh, "Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, 2017.

[22] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[23] S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. G. Moreno, B. Glocker, and D. Rueckert, "Spectral graph convolutions for population-based disease prediction," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2017, pp. 177–185.

[24] F. Alfaro-Almagro, P. McCarthy, S. Afyouni, J. L. Andersson, M. Bastiani, K. L. Miller, T. E. Nichols, and S. M. Smith, "Confound modelling in uk biobank brain imaging," *NeuroImage*, p. 117002, 2020.

[25] J. Barnes, G. R. Ridgway, J. Bartlett, S. M. Henley, M. Lehmann, N. Hobbs, M. J. Clarkson, D. G. MacManus, S. Ourselin, and N. C. Fox, "Head size, age and gender adjustment in mri studies: a necessary nuisance?" *Neuroimage*, vol. 53, no. 4, pp. 1244–1255, 2010.

[26] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[27] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J.-P. Thiran, R. Meuli, and P. Hagmann, "Predicting human resting-state functional connectivity from structural connectivity," *Proceedings of the National Academy of Sciences*, vol. 106, no. 6, pp. 2035–2040, 2009.

[28] M. D. Greicius, K. Supekar, V. Menon, and R. F. Dougherty, "Resting-state functional connectivity reflects structural connectivity in the default mode network," *Cerebral cortex*, vol. 19, no. 1, pp. 72–78, 2009.

[29] R. Kumar, L. Tan, A. Kriegstein, A. Lithen, J. R. Polimeni, H. H. Strey, and L. R. Mujica-Parodi, "Ground-truth resting-state signal provides data-driven estimation and correction for scanner distortion of fmri time-series dynamics," *arXiv preprint arXiv:2004.06760*, 2020.