Genetic Architectures of Medical Images Revealed by Registration of Multiple Modalities

Sam Freesun Friedman[1] [iD], Gemma Elyse Moran[2], Marianne Rakic[3] and Anthony Phillipakis[1]

[1]Data Sciences Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Statistics Department, Rutgers University, New Brunswick, NJ, USA. [3]Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA.

ABSTRACT: The advent of biobanks with vast quantities of medical imaging and paired genetic measurements creates huge opportunities for a new generation of genotype–phenotype association studies. However, disentangling biological signals from the many sources of bias and artifacts remains difficult. Using diverse medical images and time-series (ie, magnetic resonance imagings [MRIs], electrocardiograms [ECGs], and dual-energy X-ray absorptiometries [DXAs]), we show how registration, both spatial and temporal, guided by domain knowledge or learned *de novo*, helps uncover biological information. A multimodal autoencoder comparison framework quantifies and characterizes how registration affects the representations that unsupervised and self-supervised encoders learn. In this study we (1) train autoencoders before and after registration with nine diverse types of medical image, (2) demonstrate how neural network-based methods (VoxelMorph, DeepCycle, and DropFuse) can effectively learn registrations allowing for more flexible and efficient processing than is possible with hand-crafted registration techniques, and (3) conduct exhaustive phenotypic screening, comprised of millions of statistical tests, to quantify how registration affects the generalizability of learned representations. Genome- and phenome-wide association studies (GWAS and PheWAS) uncover significantly more associations with registered modality representations than with equivalently trained and sized representations learned from native coordinate spaces. Specifically, registered PheWAS yielded 61 more disease associations for ECGs, 53 more disease associations for cardiac MRIs, and 10 more disease associations for brain MRIs. Registration also yields significant increases in the coefficient of determination when regressing continuous phenotypes (eg, 0.36 ± 0.01 with ECGs and 0.11 ± 0.02 for DXA scans). Our findings reveal the crucial role registration plays in enhancing the characterization of physiological states across a broad range of medical imaging data types. Importantly, this finding extends to more flexible types of registration, such as the cross-modal and the circular mapping methods presented here.

KEYWORDS: Genetics, GWAS, PheWAS, autoencoders, multimodal, medical imaging, registration

Introduction

The concept of registration—mathematical maps between coordinate systems—has ancient roots, dating back at least as far as ancient Egypt.[1] There, land was surveyed with regularly knotted ropes that were gathered together—physically implementing isomorphic scaling.[2] In the medical imaging domain, registration has long been used to align different individuals to a common prototypical example of the data (ie, an atlas), or an idealized parameterization of an anatomical feature. In general, registration helps models target biological variance of interest. Variation from non-biological sources is common in medical images: for example, magnetic resonance imagings (MRIs) encode many technical and environmental artifacts, such as operator dependence or patient position in a scanner. In the supervised learning setting, there is a large literature on how to adjust for such nuisance variation. These approaches include: leveraging data augmentations,[3] adjustment based on an assumed causal model,[4] adjustment using auxiliary labels,[5] and regularizing prediction functions across different domains.[6] The unsupervised learning setting, meanwhile, presents a greater challenge as there is no outcome variable to help focus the model on relevant signals.

To ameliorate this, we demonstrate how registration affects the representations that deep neural networks learn, vastly increasing the strength of genetic, diagnostic, and phenotypic associations. This is shown across a broad range of registration methods from hand-crafted templates to nonlinear deformation fields, and a correspondingly broad range of medical data from the electro-temporal waveforms of electrocardiograms (ECGs), to two-dimensional (2D) images such as dual-energy X-ray absorptiometries (DXAs) and brain MRI slices, as well as three-dimensional (3D) spatio-temporal cardiac MRI movies.

The recent availability of large-scale multimodal measurements in biobanks[7,8] provides an opportunity to systematically study how registration affects representations of physiology. Because biobanks often assess individuals with many different modalities they allow us to compare and contrast different image types.[9] Leveraging multimodal data from the UK Biobank, this article makes the following contributions (1)
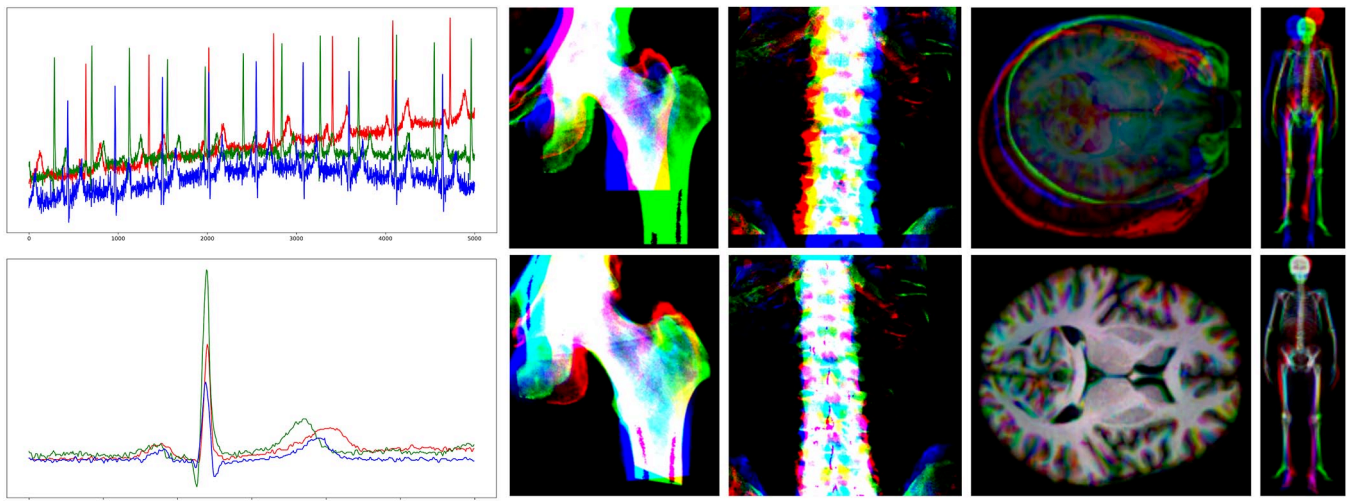
Figure 1. Examples of three individuals overlaid as red, green, and blue color channels before and after registration. The top row shows the original modality and the bottom row shows the registered version. From left to right, the modalities are the resting ECG, DXA 5 (hip), DXA 2 (lumbar spine), T1 brain MRI, and DXA 11 (whole-body skeletal).

quantify the benefit of registration across a broad range of medical image modalities, (2) show that several neural network-based methods (VoxelMorph,[10] DeepCycle,[11] and DropFuse)[12] can effectively learn registrations allowing for more flexible and efficient processing than is possible with hand-crafted or atlas-based registrations, and (3) demonstrate the breadth of downstream phenotypic analyses enriched by registration, scaling to thousands of statistical tests that comprise a phenome-wide association study (PheWAS) and the millions that constitute a genome-wide association study (GWAS).

Anatomical registrations

Many anatomical features have been used for registration. For example, 3D anatomical atlases of the brain are used to register brain MRIs.[13] Likewise, the cardiac cycle of a single heartbeat can serve as a template for both ECGs and cardiac MRIs.[14] For example, the full 10 s of the resting ECG can be registered by template matching the QRS complex, followed by alignment, scaling, and median computation.[15,16] The DXA scans can be registered with rigid homeomorphic mappings to an exemplar individual.[17] Figure 1 shows visual examples of anatomical registration by template matching and atlas alignment in five different modalities. The top row shows each modality from three different individuals overlaid before registration, while the bottom row shows the modalities overlaid after registration.

Learned registrations

Classical registration methods solve a new optimization problem for every pair, and they are therefore computationally expensive.[18-21] Recent deep learning methods propose to learn the alignment between the medical image and reference

instead[10,22-24]; supervised methods train the network and compare the output with pre-computed alignments, while unsupervised methods train their networks by learning a transformation from the image to reference, applying it and then comparing how well the aligned image matches the reference.[25] VoxelMorph, for instance, uses convolutions and spatial transformations together with a UNet architecture to learn a deformation field for each image/reference pair. It learns amortized registration by jointly maximizing agreement between the aligned image and the reference, with a transformation smoothness regularizer.[10]

While deformation fields can preserve or even increase parameterization of a modality, many registration techniques reduce dimensionality. Taking reduction to its logical extreme, DeepCycle uses a single-parameter autoencoder. This one-dimensional latent space is encoded with the inductive bias of periodicity, registering single-cell RNA expression data to the mitotic cell cycle.[11] Multimodal fusion methods, such as DropFuse, use contrastive cross-modal learning to register different modalities into a 256-dimension latent space, also greatly reducing data size.[12] Similar strategies have been pursued in a line of recent works where contrastive encoders learn joint representations of multimodal data such as natural images and their captions,[26,27] or paired clinical measurements.[28,29] Figure 2 provides a graphical summary of the learned registration techniques.

Genetic analysis of medical images

The advent of large genotyped Biobanks enables whole-genome association testing with traits derived from medical images such as MRIs, retinal images, and ECGs.[30-32] Active research has extended these genomic analyses from traits to spaces, performing the association tests in unsupervised ways.
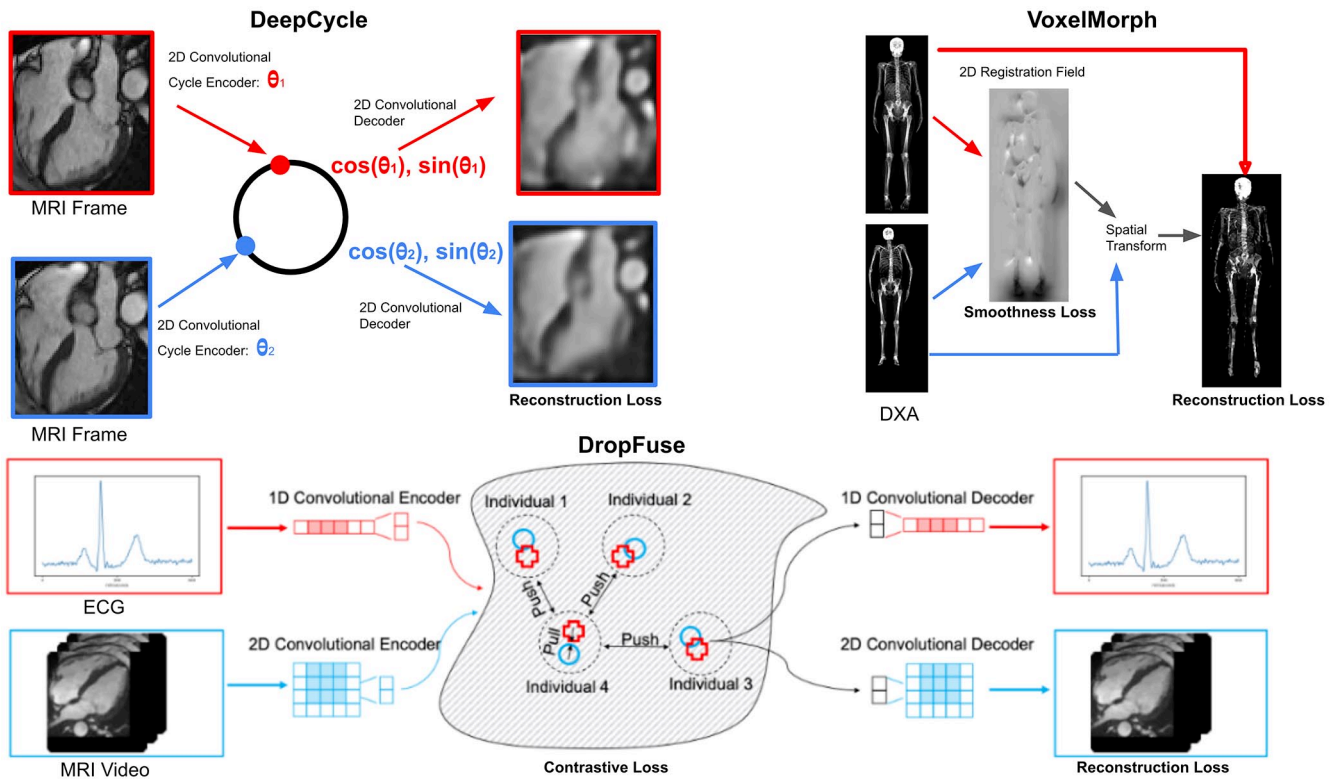
Figure 2. Three deep-learning methods used to register medical images. DeepCycle registers MRI frames to the cardiac cycle by encoding each from with a single parameter, θ, VoxelMorph learns spatial warps via pairwise amortized registration while retaining overall image dimensions and DropFuse uses dropout and cross-modal fusion to register multiple modalities together into a 256-dimensional latent space.

For example, GWAS of each dimension in a variational autoencoder, or the principal components of that encoding or even with ECG voltages from medians over every millisecond of input.[33-36] Building on this work, we demonstrate how registration uniformly results in richer phenotypic and genotypic associations over a diverse set of methods and modalities.

Methods

We train modality-specific DenseNet-style[37] convolutional encoders and decoders to reconstruct both registered and unregistered medical images from latent space bottlenecks. Table 1 details the registration techniques considered and the modalities involved. Many implementations of registration are considered, including registration learned from scratch, optimized from a predefined type of mathematical transformation (eg, homeomorphic or warp fields), or algorithmically hard-coded. The learned registrations map between individuals via deformation fields (VoxelMorph), with the inductive bias of periodicity (DeepCycle) or contrastively across modality (DropFuse). These previously described models are briefly summarized below. Code for these experiments is available in the Broad Institute's ML4H github repository: https://github.com/broadinstitute/ml4h/tree/master/model_zoo/registration_reveals_genetics.

DeepCycle

DeepCycle learns to encode data using a single-parameter latent space registered to the unit circle, demonstrating the extreme reductions possible with registration. A convolutional encoder learns a single parameter bottleneck, θ, which is registered onto the unit circle by computing (cosine(θ), sine(θ)). A convolutional decoder then reconstructs the full size input image. This model is only trained to minimize mean-squared reconstruction error. This vastly under-parameterized representation still can generate high-fidelity reconstructions, as well as generalizable and biologically informative representations, see Supplementary Video 1.

VoxelMorph

VoxelMorph is trained to minimize both smoothness and similarity losses. The smoothness loss encourages anatomical plausibility while the similarity loss ensures the fidelity of the learned registration. We trained VoxelMorph with four-chamber long axis cardiac MRI (cMRI) cine series and smoothness loss weight of 0.5. For purposes of comparison, the individual with median body mass index (BMI) was selected as an exemplar and all cMRI movies were VoxelMorphed to them.

Table 1. Modalities and registrations.

MODALITY	N	ORIGINAL SHAPE	REGISTERED SHAPE	REDUCTION	REGISTRATION METHOD	PARAMETER TUNING
ECG	42k	5000, 12	600, 12	8×	QRS Peak Align[15]	Hard-coded
DXA 11	39k	928, 352	928, 352	1×	Homeomorphic[38]	Optimized
DXA 11	39k	928, 352	928, 352	1×	VoxelMorph[10]	Learned
Brain MRI	44k	216, 256, 216	182, 218,182	1.2×	FreeSurfer[39]	Optimized
Cardiac MRI	45k	96, 96, 50	96, 96, 50	1×	VoxelMorph[10]	Learned
Cardiac MRI	45k	96, 96, 50	50	9000×	DeepCycle[11]	Learned
Cardiac MRI + ECG	38k	96, 96, 50 600, 12	256	1800× 28×	DropFuse[12]	Learned
DXA 11 + DXA 12	39k	928, 352 928, 352	256	135×	DropFuse[12]	Learned
DXA 2 + DXA 5	39k	768, 768 768, 768	256	2300×	DropFuse[12]	Learned

DXA, dual-energy X-ray absorptiometry; ECG, electrocardiogram; MRI, magnetic resonance imaging.
Each modality, its original shape, shape after registration, the resulting reduction in dimensionality, the method used to register the modality, and the way the parameters of the registering transformation are derived (ie, hard-coded with domain knowledge, optimized parameters of a known transformation or learned a new transformation via neural net approximation).

DropFuse

DropFuse is cross-modal autoencoder trained to minimize a reconstruction loss combined with a contrastive loss, which ensures that paired ECG and MRI samples are mapped to nearby points in the latent space, while discordant modality pairs are pushed away. The embeddings from each modality are fused with random dropout at each latent space coordinate. The model is trained with ECG, MRI, and DXA series pairs. The encoders for each modality are serialized separately so that inference requires only one modality to be available.

Results

To quantify the overall biological signal captured by a representation, we aggregated a broad range of phenotypes of clear biological import, including age, sex, BMI, heart rate, disease diagnoses, and principal components of genetic ancestry, see Supplementary Figures 1 to 3 for the complete lists of phenotypes used with each modality. Importantly, the phenotypes considered include both general biological features such as age and sex, along with more modality-specific phenotypes, such as the QT interval of the ECG. We then build linear probes to detect how much each representation has learned about each phenotype. It has been shown that linear separability increases monotonically as we probe deeper into the model[40]; thus by analyzing the deepest bottleneck layer, we get the best estimate for how much each phenotype is recoverable from each latent space. Both continuous and categorical phenotypes are considered. The linear probes are fivefold cross-validated to bootstrap confidence intervals on the linear and logistic regression models trained on representations of each modality before and after registration. These results are summarized in Table 2.

Improvements in the area under the ROC curve and the coefficient of determination, R^2, are shown for all of the modalities and all of the registration techniques. Many of these tasks are quite difficult, for instance diabetes diagnosis from an ECG. This is not the typical way diabetes is diagnosed or even part of typical workup. The fact that there is some discriminative power for these tasks is interesting, but the larger message of Table 2 is that the relative performance on these difficult tasks is much higher for modalities after registration.

Notably, even lossy registrations can yield more biological signals, for example from the 5000 time points of the 10s resting ECG to the 600 time points in the registered median waveform. Although much lower in overall dimensionality, registration greatly reduces intra-individual variations such as phase and baseline drift. This frees up the latent space to "spend" its expressive capacity representing more population variation, which powers downstream analyses.

Registered modalities reveal more genetic variation

The prediction of principal components of ancestry described above indicated that registration often resulted in stronger genomic association. To precisely pinpoint the genetic loci involved, we perform association tests between the latent spaces and millions of single-nucleotide polymorphisms (SNPs) throughout the entire genome. This "unsupervised" GWAS works directly on the autoencoder representations. Specifically, for each SNP, we check if the latent space centroids of the three

Table 2. Registered modalities capture more phenotypic and genetic associations.

MODALITY	REGISTERED ROC AUC	NATIVE ROC AUC	REGISTERED R^2	NATIVE R^2	REGISTRATION
ECG	0.670 (0.644, 0.696)	0.551 (0.548, 0.553)	0.412 (0.405, 0.420)	0.043 (0.035, 0.051)	QRS Peak Align[15]
bMRI	0.702 (0.699, 0.706)	0.690 (0.686, 0.694)	0.196 (0.169, 0.222)	0.156 (0.147, 0.166)	FreeSurfer[39]
cMRI	0.571 (0.567, 0.575)	0.562 (0.560, 0.564)	0.210 (0.198, 0.221)	0.048 (-0.047, 0.142)	DeepCycle[11]
cMRI	0.619 (0.612, 0.626)	0.593 (0.587, 0.598)	0.416 (0.409, 0.423)	0.284 (0.254, 0.315)	VoxelMorph[12]
cMRI ECG	0.631 (0.630, 0.633) 0.672 (0.670, 0.675)	0.593 (0.587, 0.598) 0.551 (0.548, 0.553)	0.468 (0.457, 0.479) 0.355 (0.346, 0.365)	0.284 (0.254, 0.315) 0.043 (0.035, 0.051)	DropFuse[12]
DXA 2 DXA 5	0.640 (0.637, 0.643) 0.640 (0.637, 0.643)	0.620 (0.617, 0.622) 0.626 (0.623, 0.628)	0.126 (0.121, 0.131) 0.105 (0.099, 0.111)	0.080 (0.073, 0.086) 0.065 (0.057, 0.074)	DropFuse[12]
DXA 8 DXA 11	0.687 (0.684, 0.690) 0.706 (0.703, 0.710)	0.662 (0.657, 0.666) 0.706 (0.698, 0.713)	0.144 (0.137, 0.151) 0.201 (0.194, 0.209)	0.036 (0.025, 0.047) 0.153 (0.148, 0.158)	DropFuse[12]

AUC, area under the curve; bMRI, brain MRI; cMRI, cardiac MRI; DXA, dual-energy X-ray absorptiometry; ECG, electrocardiogram; MRI, magnetic resonance imaging; ROC, receiver operating characteristic.
ROC AUC and coefficient of determination are averaged across fivefold splits of a broad range of modality-relevant tasks spanning phenotypes, diagnostics, and components of genetic ancestry. See Supplementary Figures 1 to 3 for performance on each of the tasks aggregated here.
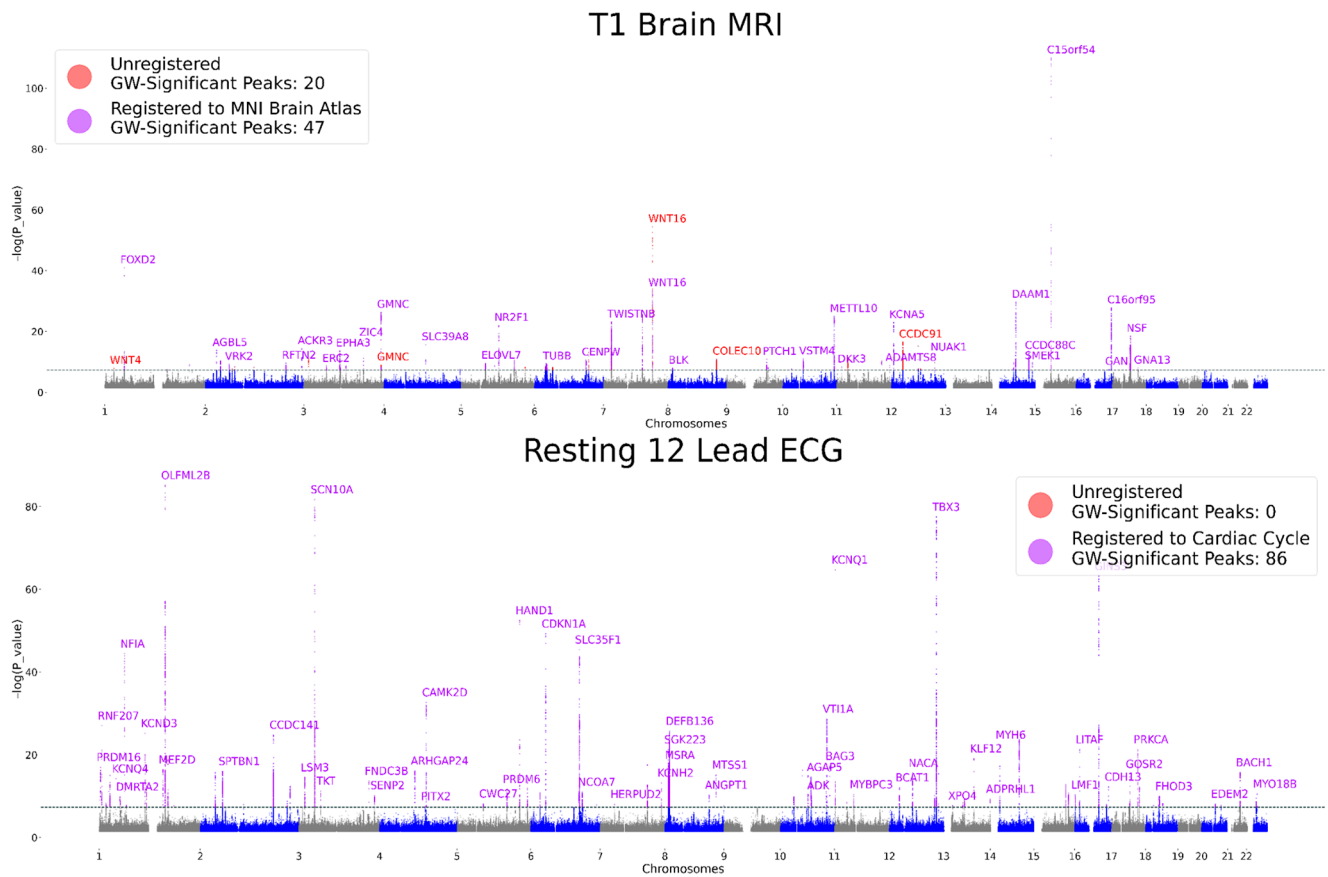

Figure 3. Manhattan plots of the T1 brain MRI (top) and the resting 12-lead ECG (bottom) with unregistered representations (lead SNPs shown in red) and with registered representations (lead SNPs shown in purple). For *P*-values and exact loci, see Supplementary Tables 1 to 3.

diploid genotypes (homozygous variant, heterozygous, and homozygous reference) are separable, as quantified by multivariate analysis of variance (MANOVA). Prior to performing MANOVA across the sets, we account for confounders from population stratification and batch effects, by removing these

sensitive features from the latent space with iterated nullspace projection.[41]

The GWAS results are shown with Manhattan plots from registered and unregistered modalities superimposed in Figure 3. In general, registered modalities have both more and stronger
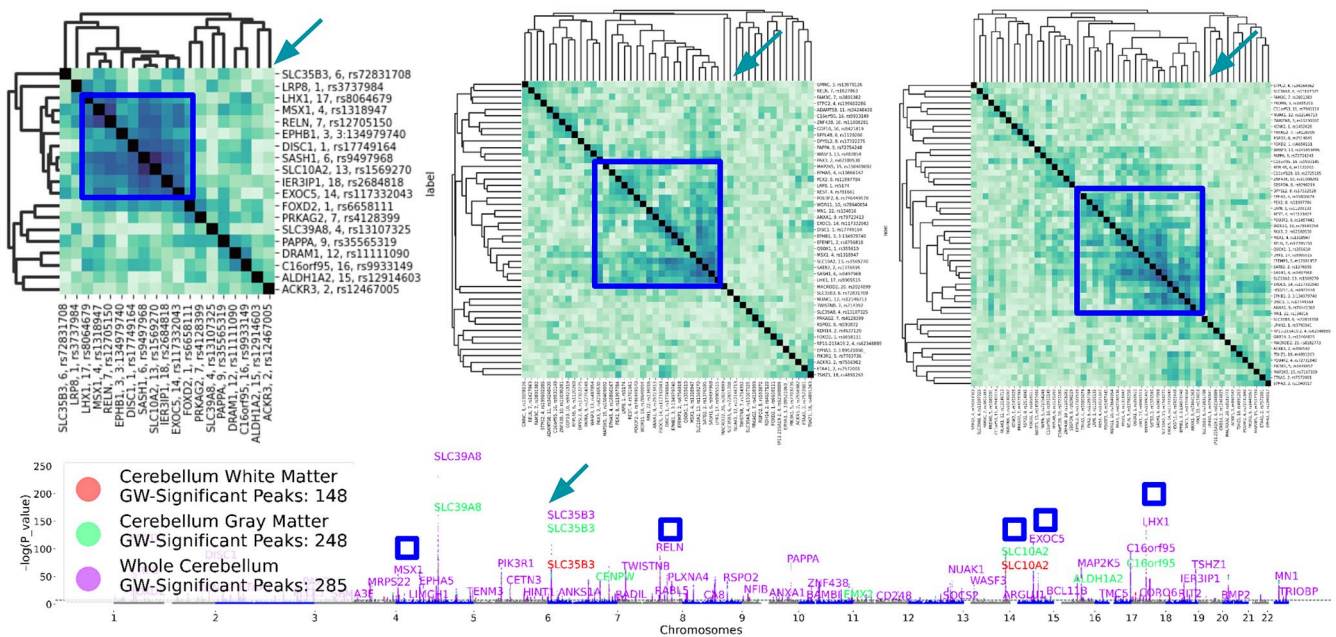
Figure 4. At top left is clustering of SNPs from autoencoders trained only on the white matter of the cerebellum, in the middle from latent spaces using only the gray matter of the cerebellum, and at top right from models trained on the whole cerebellum. Bottom panel shows Manhattan plots of the three cerebellar latent spaces with distinct but overlapping architecture.

associations. One notable exception for the brain MRIs is the gene *WNT16*, which is a site previously associated with bone density, height and body plan, not specific to the brain. In contrast, the top peak in the registered GWAS, labeled c15orf54 (chromosome 15 open reading frame 54), has previously been associated with brain region volumes and cortical structure, and this site is not significant in the native coordinate GWAS.[42,43] Similarly, the loci in genes *FOXD2*, *GMNC*, *DAAM1*, and *PTCH1* all have previously reported associations with brain-specific traits including white matter microstructure, cortical surface areas, and Alzheimer's disease biomarkers,[44-46] see Supplementary Table 1 for all T1 brain MRI lead SNPs, LocusZoom for the full GWAS summary statistics: https://my.locuszoom.org/gwas/761153/.

The GWAS of the median-waveform registered 12-lead resting ECG revealed 86 genome-wide significant loci in contrast to 0 genome-wide loci for the full 10 s unregistered ECG. The 86 loci have many previous associations with cardiovascular and specifically electrocardiographic traits. Note the large number of ion channel genes identified, specifically the sodium channels *SCN5A* and *SCN10A* and the potassium channels *KCNQ4*, *KCND3*, *KCNH2*, and *KCNQ1*. These genes play a critical role in cardiac conduction and have been associated with many cardiovascular disorders including atrial fibrillation, Brugada syndrome, long QT syndrome, and other cardiac conduction diseases.[47-49] Besides the ion channels, many of the other loci identified with nearest genes including TTN, TBX3, PITX2 have extensive previous associations with cardiovascular traits and disorders.[33,50-54] Supplementary Table 2 contains

the complete list of resting ECG lead SNPs, and full GWAS results are available at: https://my.locuszoom.org/gwas/520108/.

"Miami" plots, which superimpose two Manhattan plots after multiplying the *y*-axis of the unregistered modality by −1 for DXA series 12 scans, are shown in Supplementary Figure 4.[55] We identified 21 genome-wide significant loci, including the genes *CPED1*, *AKAP11*, and *GDF5*, which have previous associations with bone mineral density, body height, and osteoarthritis.[56] All registered DXA 12 lead SNPs are listed in Supplementary Table 3. In contrast, the GWAS of the DXA 12 latent space learned before registration identifies just two loci. The full GWAS results for the DXA 12 modality are available at: https://my.locuszoom.org/gwas/204146/.

Clustering SNPs in latent space elucidates genetic architectures

The SNP representations can be clustered to identify genetic structure in latent spaces. In particular, we perform hierarchical clustering based on the direction from the mean embedding of the homozygous reference group to the mean embedding of the heterozygous and homozygous variant groups for each lead SNP identified in a GWAS. Figure 4 contrasts latent space GWAS on three different brain MRI representations learned from the cerebellum, using only white matter, only gray matter, and the whole cerebellum. The consistency of the SNP clustering between training runs and sub-regions of the cerebellum demonstrates the reproducibility of both the SNP findings and
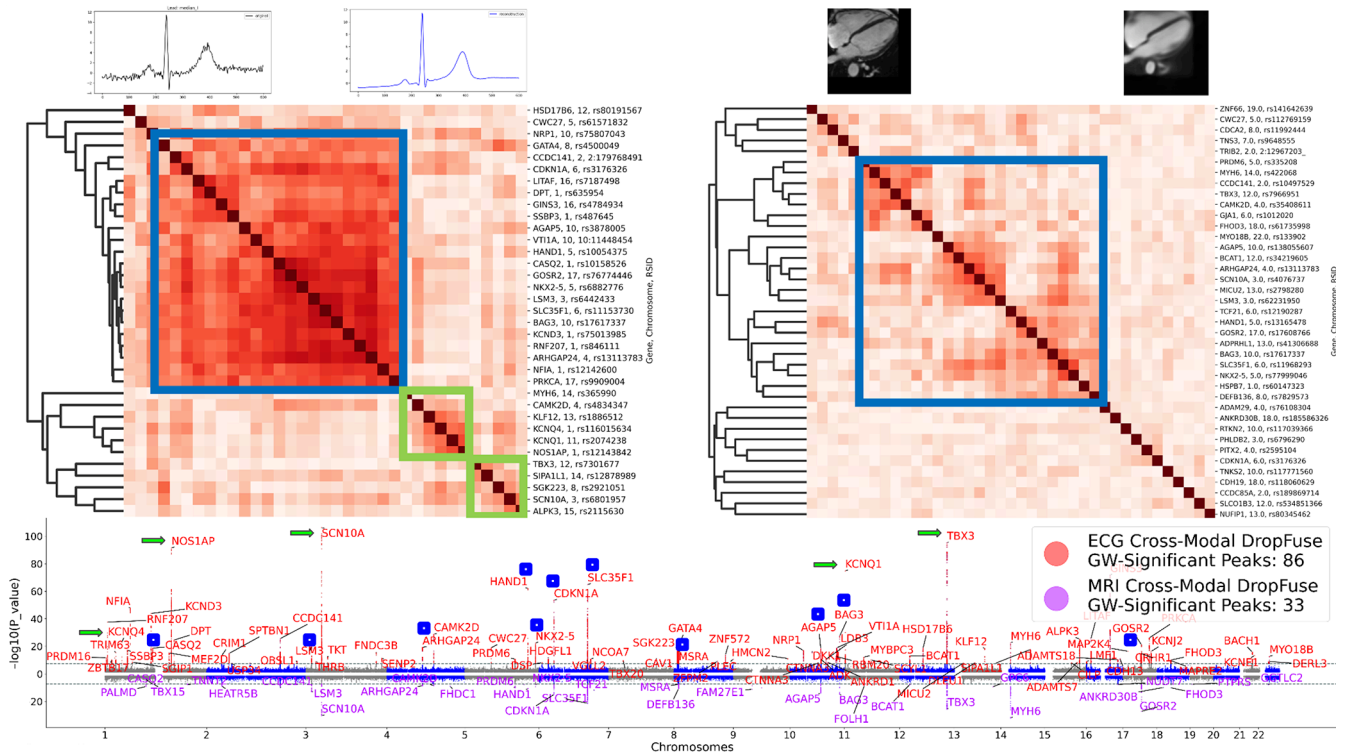
Figure 5. Hierarchical clustering of lead SNPs from the cross-modal latent space of the ECG (top left) and cardiac MRI (top right). The blue square highlights a similar clade in both models with SNPs from CASQ2, BAG3, GOSR2, NKX2-5, while the green squares highlight two clades in the ECG latent space not seen with the MRI. The bottom panel shows a Miami plot of ECG (red) and cardiac MRI (purple).

their high-dimensional latent space representations. Specifically, the blue arrows in Figure 4 highlight an SNP in the gene *SLC35B3*, which is significant in all three brain representations and consistently in an outgroup—a distal branch of the hierarchical clustering. This gene, of the Solute Carrier Family 35, member B3 has previous associations with motion sickness,[57] working memory,[58] and autosomal dominant ophthalmic outbursts.[59] In contrast to the consistent singularity of *SLC35B3*, the clustering also finds a consistent grouping of genes highlighted with blue squares in Figure 4. In that square, across all three cerebellar representations, we find the genes *LHX1*, *MSX1*, *RELN*, *EPHB1*, and *SASH1*. Compellingly, these loci all have previous associations with brain morphology[60] and cognitive traits.[61] The full GWAS results for the cerebellum are available on LocusZoom: https://my.locuszoom.org/gwas/181958/.

Hierarchical clustering in cross-modal spaces can identify modality-specific and modality-shared genetic clusters. For instance, working in the cross-modal ECG space, we find clusters corresponding to SNPs affecting the QT interval (SNPs associated with *NOS1AP* and *KCNQ1*) and SNPs related to the P-wave (SNPs associated with *SCN10A* and *ALPK3*), highlighted with green boxes in Figure 5 left. We also identify a large cluster corresponding to SNPs affecting multiple cardiac traits such as those associated with *BAG3*, *SLC35F1*, or *GOSR2*. This larger group is shared between the MRI and ECG spaces, as highlighted by the two large blue squares in Figure 5. The full GWAS results for the ECG and cardiac MRI are available at: https://my.locuszoom.org/gwas/908783/.

Registered modalities reveal more diagnostic variation

Phecodes are a taxonomy of billing codes aggregated into diagnostic labels to be more reflective of true disease phenotypes.[62] After determining phecode status from billing codes for the UK Biobank population, we conducted PheWAS before and after registration. In 50% of the subjects, we derived latent space vectors between the centroids of individuals with and without each diagnosis.[63] The remaining 50% of the individuals were embedded into the latent space and projected onto this vector. The resulting phecode vector component was tested for association with the phecode diagnosis from the EHR using a logit model corrected for age, sex, and race. After Bonferroni corrections for multiple testing, we show associations in QQ plots colored by phecode category with significant phecode associations labeled. Figure 6 shows a comparison between the PheWAS of the cross-modally registered cardiac MRI latent space and the cardiac MRI latent space trained in native coordinates. Registered latent spaces consistently identified more significant associations, specifically for cardiac MRIs 104 compared to 51 and for ECGs 63 compared to 2 (see Supplementary
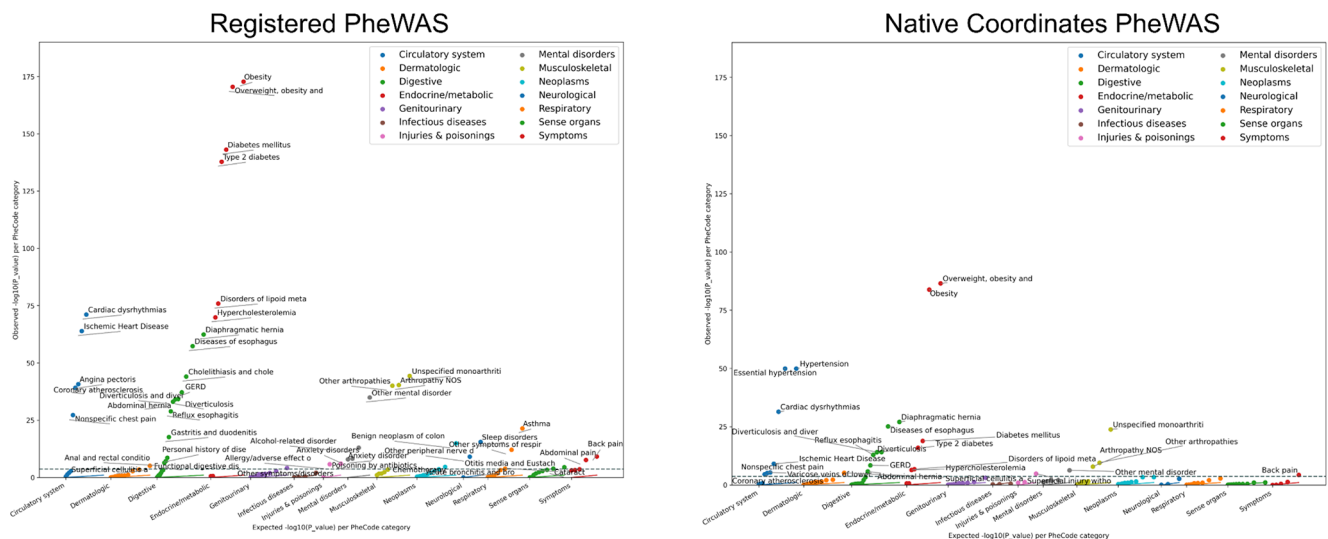
Figure 6. Phenome-wide association study QQ plots showing –log10(*P*-value) for the cardiac MRI registered by DropFuse (104 associations) and unregistered (51 associations). Phecodes are grouped and colored by diagnostic category.

Figure 5), and for brain MRIs 38 compared to 28 (see Supplementary Figure 6).

DeepCycle Groks the cardiac cycle

The most lossy registration method considered (arguably the lossiest registration possible) is with the DeepCycle convolutional encoder, which uses only a single parameter to encode the entire frame from a cardiac MRI movie. In comparison with a convolutional autoencoder without the inductive bias of periodicity, the DeepCycle MRI representations capture more biological signal, as quantified in row 3 of Table 2. Visual inspection of the DeepCycle decoder's reconstructions reveal that the parameter, θ, encodes the cardiac cycle, see Supplementary Figure 7. Without the inductive bias of periodicity, encoding instead captures the exposure value of the MRI, see Supplementary Video 1. Compellingly, the DeepCycle cardiac MRI convolutional autoencoder trained using only 50 frames spanning the heart beat from a single individual, still generalizes to the larger population. Loss curves from these models consistently exhibit the recently described Grokking phenomena.[64] Initially, training loss decreases while validation loss stagnates, but after 10 to 15 epochs validation loss also decreases, as the model learns to better "Grok" the true data distribution, as illustrated in Supplementary Figure 8. In general, registration leads to quicker and cleaner model convergence as shown by the example loss curves in Supplementary Figure 9.

Discussion

This work was motivated by the observation that autoencoder latent spaces trained from medical images in their native coordinate systems used much of their expressive power encoding

aspects of the images of limited biological significance. The top principal components in these spaced encoded information like limb orientation in DXAs or the baseline wander of ECGs. We show that registration allows models to learn representations, which use more expressive power encoding biological information such as physiological state and genetic background. This finding is consistent across diverse registration methods, imaging modalities, and organ systems.

Notably, even lossy registrations result in more biological signals. Although much lower in overall dimensionality, registration greatly reduces intra-individual variations. This frees up the latent space to "spend" its expressive capacity representing the population variation that powers the downstream biological association studies. Still, the linear probes that we used to predict phenotypes only provide a floor on how predictable a phenotype is; nonlinear reconstruction methods might be able to recover some phenotypes that linear models cannot. More expressive explainer models, such as sparse autoencoders[65] might more cleanly predict the phenotypes present in a representation. Future methods can explicitly disentangle the learned phenotypes, for instance, by factorizing modality-specific vs modality-shared signals into separate subspaces.

Limitations

There are situations where registration can introduce bias and distort associations. Anatomical atlases constructed in one population may not be appropriate for other populations with different demographics, ancestry, or disease states. An advantage of the registration learning methods (VoxelMorph, DeepCycle, and DropFuse) over template-matching methods is that they do not require prototypical individual(s) to be selected as reference. Still, learning methods are limited by the

variation present in their training data. The UK Biobank, studied here, is not representative of the world at large. Larger, more representative biobanks need to be built and existing models need to be inspected and/or corrected for potentially harmful bias. In our genetic analysis, we removed known confounds with iterative nullspace projection after model training, but this can also attenuate biological associations.

Conclusions

The best registration method for a given analysis will depend on many factors including computational resources, availability, quality and applicability of anatomical atlases, as well as the level of interpretability desired. Some registration techniques are complementary. For example, two modalities registered in space by VoxelMorph can later be cross-modally fused with DropFuse. Building up unified, biologically informative latent spaces by combining many modalities and types of registration is an exciting avenue of future research.

Acknowledgements

Author Contributions

SFF gathered the data, trained the models, created figures, tabulated the results and drafted the manuscript. SFF and AP concieved of the study. GEM identified related work and edited the manuscript. MR wrote the VoxelMorph section and guided the VoxelMorph implementation.

ORCID iD

Sam Freesun Friedman   (iD)   https://orcid.org/0000-0002-0688-2169

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online.

REFERENCES

1. Imhausen A. *Mathematics in Ancient Egypt: A Contextual History*. Princeton University Press; 2020.
2. Barnard H. Maps and mapmaking in Ancient Egypt. In: Selin H, ed. *Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures*. Springer; 2008:1273-1276.
3. Puli A, Joshi N, He H, Ranganath R. Nuisances via negativa: adjusting for spurious correlations via data augmentation. *arXiv [csLG]*. Published October 4, 2022. http://arxiv.org/abs/2210.01302
4. Wang Y, Jordan MI. Desiderata for representation learning: a causal perspective. *arXiv [statML]*. Published September 8, 2021. http://arxiv.org/abs/2109.03795
5. Makar M, Packer B, Moldovan D, Blalock D, Halpern Y, D'Amour A. Causally motivated shortcut removal using auxiliary labels. In: Camps-Valls G, Ruiz FJR, Valera I, eds. *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*. Vol 151. Proceedings of Machine Learning Research (PMLR); 2022:739-766.
6. Nguyen AT, Tran T, Gal Y, Baydin AG. Domain invariant representation learning with domain density transformations. *Adv Neural Inf Process Syst*. 2021;34:5264-5275.
7. Sudlow C, Gallacher J, Allen N, 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 Med*. 2015;12:e1001779.
8. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562:203-209.
9. Muhammed Sunnetci K, Ulukaya S, Alkan A. Periodontal bone loss detection based on hybrid deep learning and machine learning models with a user-friendly application. *Biomed Signal Process Control*. 2022;77:103844.
10. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: a learning framework for deformable medical image registration. *IEEE Trans Med Imaging*. Published online February 4, 2019. doi:10.1109/TMI.2019.2897538
11. Riba A, Oravecz A, Durik M, et al. Cell cycle gene regulation dynamics revealed by RNA velocity and deep-learning. *Nat Commun*. 2022;13:2865.
12. Radhakrishnan A, Friedman SF, Khurshid S, et al. Cross-modal autoencoder framework learns holistic representations of cardiovascular state. *Nat Commun*. 2023;14:2436.
13. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*. 2002;15:273-289.
14. de Chazal P, O'Dwyer M, Reilly RB. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Trans Biomed Eng*. 2004;51:1196-1206.
15. Carreiras C, Alves AP, Lourenço A, Canento F, Silva H, Fred A. Biosppy: biosignal processing in python. https://biosppy.readthedocs.io/en/stable/
16. Afonso VX, Tompkins WJ, Nguyen TQ, Luo S. ECG beat detection using filter banks. *IEEE Trans Biomed Eng*. 1999;46:192-202.
17. Ardeshir Goshtasby A. *2-D and 3-D Image Registration: For Medical, Remote Sensing, and Industrial Applications*. Wiley; 2005.
18. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage*. 2009;45:S61-S72.
19. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*. 2011;54:2033-2044.
20. Shen D. Image registration by local histogram matching. *Pattern Recognit*. 2007;40:1161-1172.
21. Thirion JP. Image matching as a diffusion process: an analogy with Maxwell's demons. *Med Image Anal*. 1998;2:243-260.
22. Yu EM, Wang AQ, Dalca AV, Sabuncu MR. KeyMorph: robust multi-modal affine registration via unsupervised keypoint detection. Published February 28, 2022. Accessed May 16, 2023. https://openreview.net/pdf?id=OrNzjERFybh
23. Yang X, Kwitt R, Styner M, Niethammer M. Quicksilver: fast predictive image registration—a deep learning approach. *NeuroImage*. 2017;158:378-396.
24. de Vos BD, Berendsen FF, Viergever MA, Staring M, Išgum I. End-to-end unsupervised deformable image registration with a convolutional neural network. In: Jorge Cardoso M, Arbel T, Carneiro G, et al., eds. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer International Publishing; 2017:204-212.
25. Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K. Spatial transformer networks. *Adv Neural Inf Process Syst*. 2015;28. https://proceedings.neurips.cc/paper/2015/hash/33ceb07bf4eeb3da587e268d663aba1a-Abstract.html
26. Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M. Hierarchical text-conditional image generation with CLIP latents. *arXiv [csCV]*. Published April 13, 2022. http://arxiv.org/abs/2204.06125
27. Radford A, Kim JW, Hallacy C, et al. Learning transferable visual models from natural language supervision. *arXiv [csCV]*. Published February 26, 2021. http://arxiv.org/abs/2103.00020
28. Diamant N, Reinertsen E, Song S, Aguirre AD, Stultz CM, Batra P. Patient contrastive learning: a performant, expressive, and practical approach to electrocardiogram modeling. *PLoS Comput Biol*. 2022;18:e1009862.
29. Vu YNT, Wang R, Balachandar N, Liu C, Ng AY, Rajpurkar P. MedAug: contrastive learning leveraging patient metadata improves representations for chest X-ray interpretation. In: Jung K, Yeung S, Sendak M, Sjoding M, Ranganath R, eds. Proceedings of the 6th Machine Learning for Healthcare Conference. Vol 149. Proceedings of Machine Learning Research (PMLR); 2021:755-769.
30. Elliott LT, Sharp K, Alfaro-Almagro F, et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature*. 2018;562:210-216.
31. Zekavat SM, Raghu VK, Trinder M, et al. Deep learning of the retina enables phenome- and genome-wide analyses of the microvasculature. *Circulation*. 2022;145:134-150.
32. Haas ME, Pirruccello JP, Friedman SN, et al. Machine learning enables new insights into genetic contributions to liver fat accumulation. *Cell Genom*. 2021;1:100066. doi:10.1016/j.xgen.2021.100066
33. Verweij N, Benjamins JW, Morley MP, et al. The genetic makeup of the electrocardiogram. *Cell Syst*. 2020;11:229-238.e5.

34. Yun T, Cosentino J, Behsaz B, et al. Unsupervised representation learning improves genomic discovery for lung function and respiratory disease prediction. *medRxiv*. Published online April 29, 2023. doi:10.1101/2023.04.28.23289285

35. Xie Z, Zhang T, Kim S, et al. IGWAS: image-based genome-wide association of self-supervised deep phenotyping of human medical images. *bioRxiv*. Published online May 26, 2022. doi:10.1101/2022.05.26.22275626

36. Kirchler M, Konigorski S, Norden M, et al. TransferGWAS: GWAS of images using deep transfer learning. *Bioinformatics*. 2022;38:3621-3628.

37. Iandola F, Moskewicz M, Karayev S, Girshick R, Darrell T, Keutzer K. DenseNet: implementing efficient ConvNet descriptor pyramids. *arXiv [csCV]*. Published April 7, 2014. http://arxiv.org/abs/1404.1869

38. Bradski G. The openCV library. *Dobb's J Softw Tools Prof Program*. 2000;25:120-123.

39. Fischl B. FreeSurfer. *NeuroImage*. 2012;62:774-781.

40. Alain G, Bengio Y. Understanding intermediate layers using linear classifier probes. *arXiv [statML]*. Published October 5, 2016. http://arxiv.org/abs/1610.01644

41. Ravfogel S, Elazar Y, Gonen H, Twiton M, Goldberg Y. Null it out: guarding protected attributes by iterative nullspace projection. *arXiv [csCL]*. Published April 16, 2020. http://arxiv.org/abs/2004.07667

42. Grasby KL, Jahanshad N, Painter JN, et al. The genetic architecture of the human cerebral cortex. *Science*. 2020;367:eaay6690. doi:10.1126/science.aay6690

43. Zhao B, Luo T, Li T, et al. Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat Genet*. 2019;51:1637-1644.

44. Zhao B, Zhang J, Ibrahim JG, et al. Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706). *Mol Psychiatry*. 2019;26:3943-3955.

45. van der Meer D, Frei O, Kaufmann T, et al. Understanding the genetic determinants of the brain with MOSTest. *Nat Commun*. 2020;11:3512.

46. Cruchaga C, Kauwe JSK, Harari O, et al. GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. *Neuron*. 2013;78:256-268.

47. Li W, Yin L, Shen C, Hu K, Ge J, Sun A. SCN5A variants: association with cardiac disorders. *Front Physiol*. 2018;9:1372.

48. Baltogiannis GG, Conte G, Brugada P, Sieira J, De Ferrari GM. *Sudden Cardiac Death and Channelopathies*. Frontiers Media SA; 2021.

49. Hedley PL, Jørgensen P, Schlamowitz S, et al. The genetic basis of long QT and short QT syndromes: a mutation update. *Hum Mutat*. 2009;30:1486-1511.

50. Christophersen IE, Magnani JW, Yin X, et al. Fifteen genetic loci associated with the electrocardiographic P wave. *Circ Cardiovasc Genet*. 2017;10:e001667. doi:10.1161/CIRCGENETICS.116.001667

51. Holm H, Gudbjartsson DF, Arnar DO, et al. Several common variants modulate heart rate, PR interval and QRS duration. *Nat Genet*. 2010;42:117-122.

52. Choi SH, Weng LC, Roselli C, et al. Association between titin loss-of-function variants and early-onset atrial fibrillation. *JAMA*. 2018;320:2354-2364.

53. Wang X, Khurshid S, Choi SH, et al. Genetic susceptibility to atrial fibrillation identified via deep learning of 12-lead electrocardiograms. *Circ Genom Precis Med*. 2023;16:340-349.

54. Khurshid S, Lazarte J, Pirruccello JP, et al. Clinical and genetic associations of deep learning-derived cardiac magnetic resonance-based left ventricular mass. *Nat Commun*. 2023;14:1558.

55. Paria SS, Rahman SR, Adhikari K. Fastman: a fast algorithm for visualizing GWAS results using Manhattan and Q-Q plots. *bioRxiv*. Published online April 19, 2022. doi:10.1101/2022.04.19.488738

56. Kim SK. Identification of 613 new loci associated with heel bone mineral density and a polygenic risk score for bone mineral density, osteoporosis and fracture. *PLoS ONE*. 2018;13:e0200785.

57. Hromatka BS, Tung JY, Kiefer AK, Do CB, Hinds DA, Eriksson N. Genetic variants associated with motion sickness point to roles for inner ear development, neurological processes and glucose homeostasis. *Hum Mol Genet*. 2015;24:2700-2708.

58. Donati G, Dumontheil I, Meaburn EL. Genome-wide association study of latent cognitive measures in adolescence: genetic overlap with intelligence and education. *Mind Brain Educ*. 2019;13:224-233.

59. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet*. 2016;48:709-717.

60. Smith SM, Douaud G, Chen W, et al. An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat Neurosci*. 2021;24:737-745.

61. Okbay A, Wu Y, Wang N, et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat Genet*. 2022;54:437-449.

62. Wei WQ, Bastarache LA, Carroll RJ, et al. Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLoS ONE*. 2017;12:e0175508.

63. Venn RA, Wang X, Friedman SF, et al. Deep learning of electrocardiograms enables scalable human disease profiling. *medRxiv*. Published online December 22, 2022. doi:10.1101/2022.12.21.22283757

64. Power A, Burda Y, Edwards H, Babuschkin I, Misra V. Grokking: generalization beyond overfitting on small algorithmic datasets. *arXiv [csLG]*. Published January 6, 2022. http://arxiv.org/abs/2201.02177

65. Cunningham H, Ewart A, Riggs L, Huben R, Sharkey L. Sparse autoencoders find highly interpretable features in language models. *arXiv [csLG]*. Published September 15, 2023. http://arxiv.org/abs/2309.08600