Progress and Research Priorities in Imaging Genomics for Heart and Lung Disease: Summary of an NHLBI Workshop

Matthew Nayor, MD, MPH[1]; Li Shen, PhD[2]; Gary M. Hunninghake, MD, MPH[3]; Peter Kochunov, PhD[4]; R. Graham Barr, MD, DrPH[5]; David A. Bluemke, MD, PhD[6]; Ulrich Broeckel, MD[7]; Peter Caravan, PhD[8]; Susan Cheng, MD, MPH[9]; Paul S. de Vries, PhD[10]; Udo Hoffman, MD[11]; Márton Kolossváry, MD, PhD[11]; Huiqing Li, PhD[12]; James Luo, PhD[12]; Elizabeth M. McNally, MD, PhD[13]; George Thanassoulis, MD, MSc[14]; Donna K. Arnett, MSPH, PhD[15]*; Ramachandran S. Vasan, MD[16]*

*These authors contributed equally to this work

(1) Cardiology Division, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA; (2) Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; (3) Division of Pulmonary and Critical Care Medicine, Harvard Medical School, Brigham and Women's Hospital, Boston, MA; (4) Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD; (5) Department of Medicine and Department of Epidemiology, Mailman School of Public Health, Columbia University Irving Medical Center, New York, NY; (6) Department of Radiology, University of Wisconsin-Madison School of Medicine and Public Health, Madison, WI; (7) Section of Genomic Pediatrics, Department of Pediatrics, Medicine and Physiology, Children's Research Institute and Genomic Sciences and Precision Medicine Center, Medical College of Wisconsin, Milwaukee, WI; (8) Institute for Innovation in Imaging, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA; (9) Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA; (10) Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX; (11) Department of Radiology, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts; (12) Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute, Bethesda, MD; (13) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; (14) Preventive and Genomic Cardiology, McGill University Health Center and Research Institute, Montreal, Quebec, Canada; (15) College of Public Health, University of Kentucky, Lexington KY; (16) Sections of Preventive Medicine and Epidemiology, and Cardiology, Department of Medicine, Department of Epidemiology, Boston University Schools of Medicine and Public Health, and Center for Computing and Data Sciences, Boston University, Boston, MA

Short title: Imaging genomics in heart and lung diseases

Subject terms: Genetics, genetic association studies, imaging, Computerized Tomography (CT), cardiovascular disease

Word count: 6,351

Address for correspondence:
Matthew Nayor, MD, MPH
Massachusetts General Hospital
185 Cambridge Street, CPZN #3185
Boston, MA 02114
E-mail: mnayor@mgh.harvard.edu
Telephone: 617-726-9554

ABSTRACT

Imaging genomics is a rapidly evolving field that combines state-of-the-art bioimaging with genomic information to resolve phenotypic heterogeneity associated with genomic variation, improve risk prediction, discover prevention approaches, and enable precision diagnosis and treatment. Contemporary bioimaging methods provide exceptional resolution generating discrete and quantitative high-dimensional phenotypes for genomics investigation. Despite substantial progress in combining high-dimensional bioimaging and genomic data, methods for imaging genomics are evolving. Recognizing the potential impact of imaging genomics on the study of heart and lung disease, the National Heart, Lung, and Blood Institute convened a workshop to review cutting-edge approaches and methodologies in imaging genomics studies, and to establish research priorities for future investigation. This report summarizes the presentations and discussions at the workshop. In particular, we highlight the need for increased availability of imaging genomics data in diverse populations, dedicated focus on less common conditions, and centralization of efforts around specific disease areas.

Keywords: Genomics, imaging, prevention, cardiovascular disease, metabolites, proteins

A. INTRODUCTION

The burden of heart and lung disease remains immense despite the substantial progress made by large scale population-based initiatives such as primary prevention statin therapy[1] and smoking cessation[2]. Advancing the prevention of heart and lung diseases depends on identifying individuals at increased risk and targeting lifestyle modifications or medical therapies to reduce the progression or consequences of the condition. Multivariable risk prediction models provide accurate assessments at the population level, but individual-level risk prediction remains challenging;[3] many events occur in individuals determined to be at low risk[4]. Methods that more precisely identify heart and lung disease risk and its biological determinants are needed to reduce these disorders.

Genomics assesses how variation in the genetic code contributes to phenotypic heterogeneity. In heart and lung diseases, genetic mutations have been identified to contribute to numerous (usually rare) conditions, and ever-larger genome-wide association (GWA) studies seek to associate genetic variation (common and rare) with common clinical phenotypes. Actionable insights from genetic association studies depend on mapping genetic risk variants to specific clinical phenotypes or discrete pathophysiologic mechanisms to identify true causal variants. However, the identification of true causal genetic variants is challenging with traditional GWA studies when there is high correlation among variants in close proximity to each other (linkage disequilibrium) or because a large number of variants identified by GWA studies map to intergenic areas[5]. Moreover, common heart and lung diseases are heterogeneous conditions rendering GWA studies inherently less powerful both statistically and with regard to gaining potential mechanistic insights. Precise clinical phenotyping is crucial to disentangle genetic drivers of heart and lung diseases.

Progress in bioimaging, assisted by computational advances and the use of artificial intelligence (AI) technologies, enables identification of more precise imaging features to serve as intermediate and/or advanced phenotypes for genetic analyses. The synergy between bioimaging and genomics lies in the fact that modern molecular and statistical genetics methods embrace quantitative phenotyping (i.e., description of traits on a continuous scale). Contemporary bioimaging methods provide phenotypic measurements with exceptional resolution (e.g., three-dimensional spatial sampling of 100-500 microns), and, therefore, are ideally suited as quantitative high-dimensional phenotypes. By combining state-of-the-art bioimaging techniques with genomic information, the emerging field of *imaging genomics* seeks to elucidate the genomic architecture of bioimaging traits. With higher resolution of both genomic interrogation and bioimaging phenotypes, imaging genomics has the potential to

resolve heterogeneity in the phenotypic expression associated with genomic variation, improve disease risk prediction, discover novel prevention approaches, and enable precision diagnosis and treatment (**Figure 1**).

The National Heart, Lung, and Blood Institute convened a workshop in October 2019 in recognition of this rapidly evolving field with the aim of bringing together a multi-disciplinary group of experts to review the latest cutting-edge approaches and methodologies in imaging genomics studies in heart and lung diseases, and to establish research priorities for the next phase of imaging genomics investigations[7]. In the following report, we summarize the presentations and recommendations from the workshop to provide a resource to serve as a roadmap of for the next phase of imaging genomics investigations.

B. IMAGING MODALITIES AND PHENOTYPES IN HEART AND LUNG DISEASES

Imaging genomics analyses rely on robust, reproducible, and clinically relevant discrete and quantitative imaging traits. The more precisely these imaging traits reflect underlying molecular mechanisms explained by genomic variation, the higher the expected yield of imaging genomics findings may be. A detailed review of imaging techniques themselves is beyond the scope of this workshop report. In this section, we review common imaging modalities used in heart and lung diseases with a focus on imaging techniques that are being used, or may be used in future studies, as quantitative traits for imaging genomics analysis (**Table 1**).

Imaging of the heart and vasculature

Computed tomography (CT)

CT imaging can be performed rapidly with minimal radiation exposure and provides detailed assessment of cardiovascular structures with high spatial resolution[8]. Calcium burden in the coronary arteries (coronary artery calcium, CAC), great vessels, or cardiac valves can be detected by CT imaging without intravenous contrast (or by other imaging techniques such as chest X-ray[9]) and each of these traits is associated with risk of future cardiovascular events[10,11]. Accordingly, imaging investigations have evaluated the genomic determinants of vascular and valvular calcification. Initial GWA studies for CAC identified single nucleotide variants (SNVs) that were also associated with clinically apparent coronary artery disease[12,13], lending support to the clinical (and pathogenetic) overlap of CAC and clinical atherosclerotic events. Translation of these findings to clinical risk prediction or identification of novel therapeutic targets has been limited by lack of linkage to disease-causing mechanisms[5]. Calcification of other arterial sites is also associated with cardiovascular risk. Recently, investigators identified SNVs within the

HDAC9 gene to associate with aortic calcification[14]. HDAC9 knockout mice displayed reduced aortic calcification indicating that HDAC9 may be an important regulator of vascular calcification. In addition, calcification in the aortic valve is easily obtained from CT, reproducible, and represents an earlier stage in aortic valve disease that predicts future clinical aortic stenosis[11]. GWA studies of aortic valve calcium linked SNVs in the apolipoprotein(a) gene to aortic valve calcification[11] and to clinical manifestations of aortic stenosis[15]. Clinical trials are now testing the hypothesis that pharmacologic lowering of circulating apolipoprotein(a) levels leads to reductions in aortic stenosis[16]. As shared risk factors and disease co-occurrence challenge the separation of causal factors for atherosclerotic CVD and calcific aortic stenosis in observational studies, this interventional trial may be particularly informative.

CT imaging of the coronary arteries with intravenous contrast (CT angiography) can provide detailed assessment of coronary plaque burden, location, and morphology, and the degree of blood flow obstruction as well as integrative measures such as fractional flow reserve and plaque characterization[8,17,18]. Whereas the requirement for intravenous contrast and technical expertise required for image acquisition has limited the availability of coronary CT angiography in large community-based cohorts, smaller studies have associated coronary stenoses by CT angiography with family history of coronary heart disease[19,20]. Future studies of larger samples will be important to elucidate genetic markers of relevant atherosclerotic plaque characteristics.

Echocardiography

Echocardiography provides detailed assessment of cardiac structure and function without exposure to contrast agents or radiation. As the clinical syndrome of heart failure progresses through stages of asymptomatic structural and functional cardiac remodeling, echocardiographic traits are important for imaging genomic studies of heart failure. Toward this end, GWA studies of echocardiographic traits have identified numerous SNVs, genes, and pathways linked with cardiac structural remodeling and heart failure risk[21-25].

Imaging genomics has been used to discover rare variants associated with cardiomyopathy phenotypes, such as dilated cardiomyopathy (DCM). DCM is defined as enlarged left ventricular dimensions with reduced left ventricular function. The most common underlying genetic cause of DCM is truncation variants in the *TTN* gene, which encodes the sarcomeric protein titin[26]. There is not one common variant underlying DCM but rather there are many individual *TTN* truncations that are each rare in the population. In aggregate, 1-2% of the general population carries *TTN* truncations, which are associated with subclinical

echocardiographic features, such as lower ejection fraction, and demonstrate ten-fold enrichment in DCM cases[27,28]. Importantly, not all *TTN* truncation carriers manifest with a cardiomyopathy phenotype. Taking a genotype-first approach, one study queried the electronic health record for diagnoses associated with *TTN* truncations and uncovered a higher prevalence of heart failure, atrial fibrillation, and paroxysmal non-sustained ventricular tachycardia in the absence of manifest DCM[29]. This 'genotype-first' view provides an estimate of the phenotypic expressivity of genetic variants and offers a broader assessment of the range of clinical diagnoses linked to a given genetic signature. In addition to *TTN* truncation variants, mutations in at least 50 genes lead to inherited cardiomyopathies, and each is associated with variable expressivity[30].

Future genomic studies of echocardiographic traits may use additional echocardiographic measures, such as strain imaging[31] or cardiac microstructure[32], as well as enhanced clinical phenotyping.

Cardiac magnetic resonance imaging (CMR)

CMR provides detailed structural and functional assessment of the heart and great vessels and can be performed efficiently in large community-based studies[33]. Indeed, CMR is being performed in the UK Biobank with a planned enrollment of 100,000 individuals[34]. In a report of the first ≈36,000 participants with CMR data, GWA identified 57 different genetic loci that were associated with cardiac structure and function[35]. A polygenic risk score derived from these loci was associated with dilated cardiomyopathy risk. Besides measuring heart structure and function, dedicated CMR imaging protocols can be used to evaluate specific cardiac features such as inducible myocardial ischemia[36], inflammation and fibrosis[37], and findings suggestive of specific underlying disease states[38,39]. The capability of CMR to accurately measure cardiac structure and function and complementary detailed imaging phenotypes makes it important for future imaging genomics studies.

Molecular imaging of the heart

Molecular imaging leverages targeted tracers labeled with an imaging reporter to identify specific biological processes[40]. For the assessment of coronary plaque activity, tracers can be used to assess vascular inflammation (18F-fluorodeoxyglucose with PET imaging[40,41] or ultra-small superparamagnetic particles of iron oxide imaged with CMR[42]), or microvascular calcification (representing active inflammation and higher risk for plaque rupture and identified by the 18F-fluoride PET tracer[43]). PET imaging can also be used to assess myocardial

perfusion, metabolism[44], and energetics[45]. Other imaging techniques such as magnetic resonance spectroscopy or imaging the cardiac sympathetic nervous system[46], may provide insight into disease activity and status. These emerging imaging techniques allow linkage of genetic variation with discrete biochemical processes. However, the feasibility of conducting these studies at the scale necessary to facilitate genetic interrogation is yet unproven.

Lung imaging phenotypes

Chest CT

Chest CT imaging provides excellent spatial resolution of lung architecture and diagnosis of chronic lung diseases, such as chronic obstructive pulmonary disease (COPD), emphysema, and interstitial lung disease (ILD). Qualitative chest CT assessments used in clinical practice are being expanded to allow phenotypic characterization of early disease states. Additionally, quantitative chest CT imaging approaches facilitate association with environmental exposures and genomic variation in large consortia, such as COPDGene[47], SPIROMICS[48], and the MESA Lung Study[49].

Initial imaging genomic assessments of COPD and emphysema used volumetric chest CT scans to generate measures of emphysema as defined by the percent of lung occupied by low attenuation areas (LAAs; regions with less than -910 or -950 Hounsfield Units [HU])[50]. Other quantitative chest CT measures can be used for imaging genomic analyses, including airway wall thickness and luminal diameter, small vessel pruning, large artery characteristics, gas trapping, bronchiectasis, and fibrosis, among others[47]. Subsequently, GWA studies have assessed quantitative measures of emphysema overall[51,52] (including its patterns[53,54] and distribution[55]) in conjunction with other airway phenotypic assessments[48,56,57]. While GWA studies of quantitative chest CT measures have made important discoveries, the largest COPD GWA studies have focused on spirometry, which is more readily available in large numbers of individuals. In the largest COPD GWA study to date (comprising more than 335,000 participants) 82 independent loci were associated with COPD defined by spirometry[58]. Some of these loci are more strongly associated with measures of emphysema (e.g., *FAM13A* and *HHIP*), or airways measures (e.g. *HTR4* and *RASEF*) separately, while others (e.g., *EEFSEC* and *MFAP2*) are associated broadly across phenotypes[58]. Increased sample sizes and wider availability of quantitative CT phenotypes will improve our understanding of the genetic architecture of COPD.

Qualitative and quantitative chest CT imaging assessments can also identify features associated with early forms of interstitial lung disease (ILD). Interstitial lung abnormalities (ILA;

recently formally defined by the Fleischner Society[59]) are qualitative chest CT assessments suggestive of underlying ILD or pulmonary fibrosis in individuals not previously suspected of having ILD. Efforts to identify ILD In its early stages may improve detection and prevention methods for diseases such as idiopathic pulmonary fibrosis (IPF), the most common and severe form of ILD with a mortality rate rivaling many malignancies[60]. ILAs therefore represent valuable phenotypes for imaging genomics analyses. Variants in the promoter region of the mucin 5B gene (*MUC5B*; a common variant that may explain as much as 30% of IPF risk[61]) have been consistently associated with ILAs[62-64]. A recent GWA study demonstrated four additional loci associated with both ILA and IPF (including those near genes *DDP9*, *DSP*, *FAM13A*, and *IVD*)[64]. This study also demonstrated novel genetic loci associated with ILAs but not with IPF suggesting that visually detected ILAs likely includes phenotypically heterogeneous subsets, some of which may not correlate with overt IPF[64].

Quantitative efforts to identify early stages of ILD and pulmonary fibrosis include assessments of high attenuation areas (HAAs; typically between -600 and -250 HUs), local histograms (referred to as “interstitial features”), and deep learning-based lung textural assessments of lung fibrosis[65,66]. These quantitative chest CT features can potentially detect, characterize, and grade early stages of pulmonary fibrosis. However, these measures are not interchangeable, demonstrate varying degrees of sensitivity and specificity when compared to visual assessment of pulmonary fibrosis[66,67], and can be affected by technical differences between studies.

Molecular imaging of pulmonary fibrosis

Molecular probes targeting several components of pro-fibrotic pathways are under development for imaging fibrotic activity in IPF. By discovering early activity in pulmonary fibrosis before it is apparent by CT imaging, these molecular probes may identify earlier forms of the disease prior to overt lung fibrosis and can potentially be used to assess therapeutic responses to antifibrotic agents.

Molecular probes have been developed for type I collagen, a primary component of fibrosis (^{68}Ga-CBP8)[68] and for collagen cross-linking (Gd-CHyd)[69]. In clinical studies, patients with IPF had increased lung uptake of ^{68}Ga-CBP8, but it was also found in areas of normal appearing lung suggesting that ^{68}Ga-CBP8 may be sensitive in detecting regions of active collagen deposition prior to manifest fibrosis evident on CT imaging[70]. Probes also target drivers of fibrosis, such as integrin $\alpha_v\beta_6$[71], and the C-X-C chemokine-receptor type 4 (CXCR4)[72]. Somatostatin receptor 2 is expressed on fibroblasts and increases in fibrosis[73]. Somatostatin

receptor expression in humans with pulmonary fibrosis can be assessed using ^{111}In-octreotide scintigraphy (Octreoscan) and ^{68}Ga-DOTANOC PET-CT[74,75]. Beyond fibrosis-specific pathways, broader processes such as vascular leak[76], coagulation, and inflammation have been implicated in the development of pulmonary fibrosis and each can be assessed with specific molecular probes[77-80]. Similar to molecular imaging of the heart and vasculature, molecular imaging of the lungs is not yet being done at the scale necessary for GWA studies.

C. ARTIFICIAL INTELLIGENCE FOR BIOIMAGING

Terminology and methods

AI is a general term describing disciplines that aim to mimic human intelligence[81]. Here, we focus on the relatively narrow functions of data interpretation and risk prediction, but AI technologies capable of accomplishing complex human tasks are also being developed and may have additional wide-ranging applications[82]. At the core of AI is machine learning (ML), which uses mathematical models to substitute for human thought and decision-making. ML algorithms take inputs (e.g., clinical characteristics) and map them to given outputs (e.g., mortality), optimizing the fit of the model to achieve optimal performance. ML models are not explicitly programmed to achieve a given task. Instead, they minimize the *'cost'* of the model (the proportion of misclassified patients)[81]. This is regarded as the *learning* process of ML algorithms. Accordingly, ML algorithms gradually increase their accuracy as they 'see' and 'learn' from more data.

Application of AI for identification of anatomical structures and pathologies

Deep learning (DL) is a type of ML that can use pixels or voxels from bioimaging data as inputs to *neural network* models, which iteratively relate inputs to outputs in flexible models allowing complex nonlinear functions[83]. DL is well-suited for automating time-intensive tasks in imaging interpretation[84]. To train such an algorithm, images are used as inputs, and labeled segmentations are needed as outputs. The DL algorithm can then learn to automatically segment radiological images to identify potential pathologies. For example, large contemporary datasets (such as UK Biobank) provide vast amounts of imaging data, which are mostly present in "unlabeled" formats without phenotypic measures or annotation. Several groups have recently developed DL models to derive clinically relevant phenotypic data from these images[85-87]. In addition, DL algorithms can identify diabetic retinopathy from retinal images with a higher risk of vision loss in individuals with diabetes[88]. There are several potential challenges in developing effective DL algorithms for phenotypic characterization, however. These models may not

translate well from one sample/dataset to another, as even slight differences in image acquisition and quality may negatively impact model performance[89,90]. Furthermore, DL algorithms are statistical models that do not understand the underlying meaning of the inputted information. It is therefore conceivable that decisions are being made not based upon disease-related imaging abnormalities, but based on acquisition settings or other imaging features, such as markers on chest X-rays[91]. In addition, it is increasingly recognized that DL algorithms may perpetuate health inequities based on gender, race, ethnicity, or other factors if training data sets lack demographic diversity or if outcomes partially reflect underlying healthcare disparities[92].

Identification of abnormalities and novel imaging phenotypes

AI-derived outcomes may provide novel bioimaging phenotypes that are not apparent to the human eye through unbiased assessments of imaging characteristics or imaging data. For example, DL algorithms can extract complex information from low cost and widely available images such as chest X-rays that can be used to predict incident lung cancer or cardiovascular, respiratory, or lung cancer mortality[93,94]. *Radiomics* is a *feature-generative* technique that derives new imaging measures based on mathematical properties of the imaging data including intensity distribution, texture, and shape[95]. Radiomic analysis of coronary plaques may potentially improve prediction of cardiovascular events via detailed description of high-risk properties of atherosclerotic lesions[96,97]. Analysis of radiomic features describing texture properties of the myocardium on CMR images may enable detection of subtle differences in the myocardial intensity values to better distinguish patients with subacute and chronic myocardial infarction[98]. Radiomics features may also help to differentiate pathologies with similar gross appearance to the human eye, such as hypertensive heart disease and hypertrophic cardiomyopathy[99]. It is important to note that potential association of radiomic features with disease entities does not imply causation[90]. Connecting biological processes with specific radiomic features is, therefore, only speculative at present.

AI methods are being applied to CT images to develop novel disease classifications in COPD (**Figure 2**). For example, DL algorithms have been trained to classify emphysema on lung CT and scores derived from the DL algorithms can improve classification of emphysema and outcome prediction when compared to visual scoring[101]. Using these methods, investigators identified two distinct trajectories of emphysema progression: 1) airway→tissue type in which large airway abnormalities precede emphysema; and 2) tissue→airway type in which small airway dysfunction and emphysema antedate large airway wall abnormalities[102]. Unsupervised

ML methods have also been applied to define novel emphysema sub-phenotypes based upon texture and anatomical location of emphysema on lung CT scans[100].

Radiomics and DL-based imaging traits have been mostly evaluated in small pilot studies, thus far; larger scale observational studies[103], clinical trials, or innovative study designs such as *in silico* trials, are necessary to assess their superiority over traditional imaging measures for clinical applications.

D. EMERGING GENOMIC APPROACHES

The GWA era has yielded many important mechanistic insights and information regarding genetic variants associated with higher disease risk. However, several barriers to clinical translation of genomic findings are present currently. First, there remains substantial unexplained heritability for many common complex diseases of multifactorial etiology. Second, the underlying mechanisms linking genetic variants with imaging traits, and ultimately with disease states, is currently lacking for many SNVs[5]. Third, variable expressivity (penetrance) of genetic variation needs to be better understood through analysis of gene-gene and gene-environment interactions, and the variability of such interactions in the time domain over the lifecourse. Fourth, genetic determinants of bioimaging traits may vary based on racial/ethnic background and should be studied in diverse populations. Advances in the resolution of genetic sequencing and harnessing complementary multi-omic analyses have the potential to address these current challenges thereby broadening the impact of imaging genomics investigations.

Whole genome sequencing

The initial phases of GWA studies used genotyping arrays to directly ascertain 200,000–1 million genetic variants and then leveraged reference panels to impute a larger number of variants[11,12,14,22]. However, historical reference panels have important limitations including inadequate coverage of rare or private variants, difficulty identifying causal variants due to linkage disequilibrium blocks, and the potential for bias when non-European populations are less well represented in the derivation samples.

Whole genome sequencing (WGS) provides expanded coverage of the entire genome thereby overcoming the reliance on reference panels and imputation and enabling a balanced assessment across the genome. To harness WGS for genetic discovery, the National Heart, Lung, and Blood Institute initiated the large-scale WGS TOPMed (Trans-Omics for Precision Medicine) initiative (www.nhlbiwgs.org). TOPMed was established to advance genetic discovery by addressing some of the limitations of the European-ancestry centric, and array-based GWA

studies conducted previously. Over 80 existing studies with phenotypic data are part of TOPMed. Across these studies, TOPMed has generated WGS data on over 181,000 participants, at an average sequencing depth of 38x. Ethnic and racial diversity is a priority in TOPMed with 40% participants of European ancestry, 31% of African ancestry, 15% of Hispanic ancestry, and 9% of Asian ancestry.

Data on the first 53,831 participants with available sequencing data in TOPMed were recently reported[104]. In this subsample, 410 million genetic variants were identified, of which 97% had a minor allele frequency below 1%. In fact, 46% appeared only in a single individual ('singletons'). When this dataset was used as a reference panel for genotype imputation, the variants with minor allele frequencies as low as 0.01% were imputed with high quality (r^2 >0.9) in individuals of both European and African ancestry allowing the inclusion of much rarer variants in GWA studies than was possible with previous reference panels. The inclusion of participants from multiple ancestry groups leads to a substantial reduction in the size of linkage disequilibrium blocks at most loci (i.e., trans-ancestry fine-mapping) enhancing the ability to identify causal variants within large blocks of the genome with relatively high linkage disequilibrium in certain populations[105] (NHLBI TOPMed imputation server can be found at https://imputation.biodatacatalyst.nhlbi.nih.gov/).

With resources available through TOPMed, WGS data can be combined with high quality complementary "omics" data (including epigenetics, transcripts, metabolites, and proteins) and detailed phenotypic information (available phenotypic data can be found at https://biodatacatalyst.nhlbi.nih.gov). Numerous studies included in TOPMed have heart and lung imaging phenotypes including CAC, echocardiograms, CT, and MRI. GWA studies leveraging WGS are now underway and are expected to identify variants that are not well covered by standard imputation approaches such as rare variants, variants on the X chromosome, and variants restricted to non-European populations, providing the potential for discovery of novel disease-causing variants.

Trans-omic signatures of imaging traits

High-dimensional molecular data (e.g., transcriptomics, proteomics, metabolomics) can provide complementary information to genetic studies for identifying mechanisms of disease related to imaging phenotypes. These tools can be used to better understand the downstream implications of genetic variants, to improve the precision of CVD phenotyping by defining endophenotypes that can be correlated with imaging traits, or to identify novel disease-associated biological pathways for mechanistic investigations[106]. Importantly, molecular

signatures may reflect causal links between genetic variation and clinically relevant phenotypes (such as bioimaging traits), but they also might represent compensatory mechanisms or epiphenomena associated with both the genetic variant and the phenotype, but without a causal relation. Thus, an integrative trans-omic approach that assimilates data across profiling platforms will be the ideal approach to elucidate the potential roles of different molecular signatures in disease pathogenesis. Herein, we discuss investigations using proteomic and metabolomic data to examine molecular signatures of adverse cardiac structural remodeling traits as examples of how high-dimensional data derived from different profiling methods may be used to complement imaging genomics analyses[107].

High-throughput proteomic and metabolomic data and imaging phenotypes

Numerous proteins have been associated with imaging signs of cardiac remodeling and with heart failure risk in prior studies assessing a limited set of candidate proteins[108]. Technological advances in high-throughput proteomic profiling platforms now enable deep and relatively unbiased profiling of more than 5000 of the ≈20,000 human proteins[109], which can provide a broader view of the relations of the circulating proteome with cardiac imaging traits. For example, proteomic signatures of coronary flow reserve can be identified in peripheral blood[110]. Such a molecular signature can be used as a discrete disease phenotype for other analyses (such as genetic discovery) or to uncover biological pathways relevant to heart failure pathobiology. Proteomic profiling has also been used to evaluate how inflammatory pathways may mediate the association of general comorbidity and echocardiographic traits in patients with heart failure with preserved ejection fraction[111]. A discovery platform comprising ≈1300 plasma proteins was used to identify novel associations of proteins representing different biological pathways with echocardiographic traits[112].

Hundreds of small molecule metabolites can also be measured from the blood using nuclear magnetic resonance- or mass spectrometry-based techniques[113]. Heart failure and its imaging correlates represent attractive conditions for metabolomic analysis as the failing heart undergoes extensive metabolic remodeling with reduced fatty acid oxidation and a shift to using glucose as its primary fuel source[114]. The reduction in fatty acid oxidation is reflected by increased circulating levels of long-chain acylcarnitines[115]. In individuals with aortic stenosis, long-chain acylcarnitines are present in higher levels in blood in those with adverse cardiac remodeling phenotypes, such as severe left ventricular hypertrophy and reduced LV systolic function[116]. Within 24 hours after aortic valve replacement, a drop in circulating long-chain acylcarnitine concentrations can be observed, demonstrating changes in cardiac metabolism

that precede detectable changes in cardiac structure or function[116]. Additional metabolites, such as lipid ceramides[117], kynurenine (a tryptophan derivative implicated in vasodilation), and aminoadipate are also associated with cardiac structure and function[118]. Metabolite profiling, therefore, can identify signatures in circulating blood that correspond to molecular abnormalities in the heart itself; these abnormal metabolite profiles relate to imaging measures of heart structure and function and may predict future heart failure. It can be a challenge, however, to decipher whether these molecular signatures are causal in the disease process. Leveraging longitudinal molecular data in carefully phenotyped individuals before the onset of manifest heart or lung disease will be an important tool to elucidate which molecular profiles or pathways might play causal roles.

Correlating circulating protein and metabolite concentrations with genetic determinants

Circulating protein and metabolite concentrations are partly heritable, with a relatively large amount of their variation explained by genetic factors[119-121]. Protein or metabolite levels that are related to imaging traits can be associated, therefore, with genetic variation (i.e., protein quantitative trait loci [pQTLs] or metabolomic quantitative trait loci [mQTLs]) to discover new genetic markers of disease risk[120]. The statistical technique of *Mendelian randomization* can provide data in support of a causal role for a protein or metabolite in disease pathogenesis by leveraging the random allocation of pQTLs/mQTLs among individuals[122]. Genetic determinants of protein/metabolite levels can also be used to "replicate" findings from proteomic and metabolomic analyses in external cohorts by using their quantitative trait loci as proxies for the level of the molecule itself[112,123]. This approach was recently used to *replicate* relations of novel proteins with echocardiographic traits from one community-based sample in a large, international, genetic consortium[112]. Alternatively, metabolomic and proteomic profiles can be used to better understand the molecular mediators of the associations of genetic variants with clinical traits[107,124].

E. METHODOLOGICAL DEVELOPMENTS IN IMAGING GENOMICS

Imaging genomics involves analysis of high-dimensional data sources and, therefore, poses challenges for standard data analysis and computing techniques. In the following section, we provide an overview of methodological developments in computational strategies and bioinformatics approaches aimed at harnessing massive datasets for imaging genomics analysis.

Computational strategies for heritability and genetic association studies

The standard statistical genetics model for both heritability and association studies becomes computationally impractical when applied in large-scale imaging genetic studies. Through algorithmic and hardware acceleration approaches, the computational burden can be reduced from $10^{10\text{-}12}$ to $10^{1\text{-}2}$ hours.

Algorithmic acceleration of the standard genetic model

For imaging genomics studies, a relevant model needs to be evaluated for every imaging trait and every genetic polymorphism. A typical voxel-wise MRI/CT analysis may involve several hundred thousand traits each detailing regional information. These analyses require maximizations of likelihood for each of the voxel-wise traits and can each take between 10-50 iterations. Therefore, a GWA of 1,000,000 SNVs and 100,000 voxels would require 10^{11} maximizations of likelihood. Each iteration of the likelihood algorithm requires the inversion of the covariance matrix, which is a substantial computational effort that grows as $N^{2\text{-}3}$ with increasing sample size (N) and becomes impractical for large-scale imaging genomics efforts.

Algorithmic acceleration of the standard genetic model can provide substantial improvements in the efficiency of conducting large-scale imaging genomics analyses[125-127]. This is accomplished by reducing the computational effort associated with the inversion of the covariance matrix for each likelihood calculation. The eigen value decomposition approach performs an orthogonal transformation to diagonalize the covariance matrix, thereby making the matrix inversion computationally trivial[128]. To reduce the iterative burden of maximum likelihood estimates, a two-step ordinary linear squares followed by a weighted linear squares approximation (*Fast and Powerful Heritability Inference* [FPHI]) can be used to solve the maximum likelihood estimation non-iteratively[125]. The FPHI model can then be expanded to a *Fast and Powerful Genome-wide Association* (FGPA) for genotype association analysis of the full model including measured SNVs[126]. Additional computational performance can be achieved by testing the statistical significance of association using the Wald test (FPGA-Wald). Hence, the algorithmic approximation of standard genetic models using FPHI, FPGA and FPGA-Wald, provides significant improvement in computational efficiency versus classical approaches, reducing the computational complexity from $N^{2\text{-}3}$ to N^{1}. Detailed discussion of these acceleration approaches is included in the **Supplemental Material**.

Hardware acceleration of imaging genomics computations

Even with the algorithmic approximations discussed above, the computational burden of large-scale imaging genomics studies is immense and calls for efficient implementation using modern hardware. Computational clusters are built of nodes equipped with central processing and graphics processing units (CPU/GPU) offering multiple computational cores (typically 2-64 for CPUs and 1000-8000 for GPU). The CPU and GPU version of FPHI and FPGA can be implemented using linear algebra software libraries that optimize the code for parallel scientific computing in a CPU and GPU environment (OpenMP; https://www.openmp.org). FPHI and FPGA algorithms are coded using cuBLAS (https://developer.nvidia.com/cublas) linear algebra libraries for GPU computing. More information is available in the **Supplemental Material**. All algorithmic, software, and hardware approaches discussed above are implemented in the solar-eclipse software (http://www.solar-eclipse-genetics.org/) and are freely available for download, use, and distribution.

Bioinformatics strategies for imaging genomics studies

Given the high dimensionality of both imaging and genomics data, massive univariate analysis methods (i.e., single imaging trait vs. single variants) face a huge burden for multiple statistical testing correction, which results in reduced detection power. Various bioinformatics strategies can be used to reduce the multiple hypothesis testing burden and to improve the interpretability and mechanistic insight gained from imaging genomics investigations.

Meta- and mega-analysis

Meta- and mega-analyses can be used to boost the statistical power of imaging genomics analyses by creating massive datasets. In meta-analysis, summary data from many collaborating cohorts can be combined after overcoming logistical and regulatory challenges in merging participant-level data[22]. In mega-analysis, participant-level data on many individuals are centrally analyzed.

Polygenic and multivariate analysis

Another strategy to improve detection power is to leverage associations involving multiple imaging or genomic markers, thereby reducing the effective number of tests performed. By combining SNVs or imaging traits into statistically or biologically related groups, these techniques may also help deconvolute mechanistic complexity and lead to better understanding of disease subtypes.

Polygenic risk scores (PRS) can capture the accumulated effect from a set of trait-related SNVs[129]. While each SNV may not have a statistically significant effect on the trait, the collective effect captured by the PRS may explain a considerable portion of the trait variance. By examining a PRS instead of many individual SNVs, the burden for multiple hypothesis testing correction is reduced. Other methods can be used to investigate associations involving multiple genomic markers, such as multiple regression to examine the joint effect of several SNVs (chosen based on prior knowledge) on each imaging trait[130]. Gene-based GWA can estimate the aggregate effect of all the SNVs within each gene on each imaging trait[131]. Commonly used techniques for multi-trait analysis include: 1) combining single-trait results (e.g., selecting the SNV with the minimum p-value[132]); 2) applying univariate analysis on a small number of extracted trait features (e.g., the average trait or first few components from principal components analysis[133]); and 3) classical multivariate analysis methods such as multivariate analysis of variance (MANOVA[134]) and generalized least squares regression[135].

Pathway and network enrichment analysis

Enrichment analysis organizes data into sets based on functional pathways and networks, thereby providing natural connections to biological mechanisms. There are two standard types of methods for enrichment analysis: 1) *threshold-based* methods (e.g., hypergeometric test or Fisher's exact test) are designed to identify pathways or networks over-represented by GWA hits; and 2) *rank-based* methods (e.g., GSEA-SNP[136]), which employ a Kolmogorov-Smirnov-like running sum to quantify the degree to which a gene set is over-represented at the top of the gene list ranked by the GWA results. Similarly, network-based GWA studies can identify functional modules from biological networks that are enriched for top GWA findings[137]. Most existing methods analyze tissue-free networks without reflecting phenotypic specificity, but novel tissue-specific functional interaction networks are being developed[138].

Regularized regression and correlation analyses

Regularized regression models[139,140] and bi-multivariate correlation models, such as sparse canonical correlation analysis (SCCA)[141,142], can be used to examine multi-SNV-multi-trait associations. These models often include a sparsity-inducing regularization term to identify a smaller number of relevant imaging and genomic markers and to minimize overfitting[139,140]. Regularized regression and SCCA models benefit from relative ease of interpretation.

Outcome prediction and joint association learning and outcome prediction

Another challenge in imaging genomics is integration of imaging and genomics data for improved prediction of clinical outcomes. The most straightforward methods are to concatenate imaging and genomic data together, and then apply conventional predictive models[143]. Advanced ML and DL methods are also being developed to co-relate genetic and imaging markers with relevant outcomes[144,145]. ML models can also be used for joint exploration of the associations among genomics, imaging traits, and outcome[146].

F. RESEARCH OPPORTUNITIES AND PRIORITIES (Summarized in **Table 2)**

Increased availability of imaging genomics data in diverse populations

While consortia and large biobanks with imaging and genomic data are proliferating, initial datasets mostly comprised white individuals of European descent. Multiethnic large samples to capture the diversity and spectrum of phenotypic variation are required for the benefits of imaging genomic findings to be shared equally across populations. TOPMed and the Million Veterans Program[147] have both sought to enroll racially diverse samples and efforts to leverage and expand the available imaging phenotypes within these initiatives and in new cohorts should maintain the important focus on establishing representative study samples with diversity of race, ethnicity, age, and sex, among other factors.

Data harmonization and infrastructure

The unprecedented scale, complexity, and heterogeneity of multidimensional big datasets available to the research community (including genetics, multi-omics, bioimaging, electronic health record data) requires effective data integration and harmonization methods to realize the full potential of these invaluable resources. Tools are being developed in partnerships with NIH Data Commons[148] and NHLBI BioData Catalyst[149] to leverage cloud-computing resources for sharing and analyzing TOPMed and imaging datasets within a secure cloud environment. Unifying imaging biomarkers and genomic data requires the availability of imaging data in formats that can be shared efficiently and securely. These efforts would, therefore, benefit from efficient and scalable computational infrastructure to facilitate different types of large-scale collaborations across cohorts (e.g., cloud computing for teams working with one single centralized data repository) and federated learning for teams working with distributed data sets.

Centralization of efforts around specific disease areas

Translating imaging genomics findings from the initial discovery of SNV-imaging trait associations to the use for clinical risk prediction or mechanistic insight requires coordinated efforts across disciplines. This may require large, diverse datasets for the initial discovery of genetic variants, replication in other populations, complementary molecular phenotyping through gene expression, metabolite profiling, or proteomic investigations, and functional and mechanistic studies harnessing robust model systems. Integration of these complex steps would be facilitated by centralized "nodes" of interdisciplinary investigators centered around disease areas or pathophysiologic processes (e.g., inflammation or fibrosis). These centralized efforts will also facilitate the development of standards for imaging genomics methods and related interpretations norms.

Methodological advances in imaging genomics

Despite the tremendous progress made in recent years in the field of imaging genomics and the development of necessary tools for its implementation, substantial methodological challenges remain. Statistical and computational tools to integrate these high-dimensional data sources are available, but they are variable across different fields and standardized methods and interpretation norms are not yet established. These analyses have been largely conducted to discover statistical associations; leveraging longitudinal measures and more advanced statistical techniques may enable causal inferences regarding the genetic etiologies of imaging traits.

Dedicated focus on less common conditions

Most consortia for imaging genomics analyses have been formed to enhance statistical power to evaluate common variants with relatively low effect sizes. While efforts to better characterize and understand genetic underpinnings of common disease will undoubtedly continue, consortia should also be formed to study genetic correlates of imaging traits in less common diseases. In many cases, while the manifest disease itself may be relatively rare, imaging features of its pathogenesis can be detected years earlier and in a broader population (such as aortic calcification for aortic stenosis or ILAs for IPF). By centralizing efforts to study less common diseases, investigators may identify pathophysiological mechanisms that are applicable to other disease forms as well.

F. SUMMARY

The emerging field of imaging genomics seeks to combine state-of-the-art imaging techniques with genomic information to understand the genomic architecture of bioimaging traits. Several common heart and lung diseases (e.g., atherosclerosis, COPD) are caused by a combination of environmental and genetic factors, rendering imaging genomics an especially powerful tool for improving the precision with which they are characterized across the lifecourse. Measuring precise bioimaging phenotypes with relevance to clinical disease and elucidating their genomic contributors can improve understanding of disease pathogenesis, identify relevant biological pathways, lead to the discovery of novel therapeutic targets, and facilitate disease prevention characterized by more precise risk prediction and screening and enhanced opportunities for prevention of clinically meaningful events.

Disclaimer: Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect the official views of the NHLBI, the National Institutes of Health, or the US Department of Health and Human Services.

Acknowledgments: None

Sources of funding

Dr. Nayor acknowledges support from grant K23-HL138260 from the National Heart, Lung, and Blood Institute. Dr. Shen acknowledges support from NIH grants R01 EB022574, R01 LM013463 and U01 AG068057 and NSF grant IIS1837964. Dr. Hunninghake acknowledges support from the National Heart, Lung, and Blood Institute and NIH with grants R01 HL 111024, R01 HL135142, and R01 HL130974. Dr. Kochunov acknowledges support from NIH grants R01 EB015611, P50 MH103222, R01 NS114628 and U01 MH108148. Dr. Barr acknowledges support from the National Heart, Lung, and Blood Institute with grants R01 HL077612, R01 HL093081, R01 HL121270 and R01 HL1422028. Dr. Broeckel acknowledges support from NHLBI through the grants 1R01HL125580, 1R01HL140493 and 1R01HL107577. Dr. Caravan acknowledges support from the NHLBI with grants R01 HL109448, R01 HL116315, R01 HL131907, R33 HL154125. Dr. Cheng acknowledges support from the NHLBI with grants R01 HL134168, R01 HL131532, R01 HL143227, R01 HL142983. Dr. de Vries was supported by National Heart, Lung, and Blood Institute grants R01HL146860 and American Heart Association grant 18CDA34110116. Dr. McNally was supported by NIH U01 HG011169, NIH U01 HL131914, NIH R01 HL128075 and the American Heart Association SFRN on Arrhythmias and Sudden Cardiac Death. Dr. Arnett acknowledges support from the National Heart, Lung, and Blood Institute with grants R01 HL55673. Dr. Thanassoulis is supported by the Canadian Institutes of Health Research, the Fonds de Recherche Québec- Santé, and by NHLBI grant R01 HL128550. Dr. Vasan acknowledges support from the Framingham Heart Study (FHS) Contracts NO1-HC-25195, HHSN268201500001I and 75N92019D00031 from the National Heart, Lung, and Blood Institute and by NIH grants HL107385, HL126136, HL93328, HL 142983, HL143227 and HL 131532, U01HL146382, R01HL146860 and R01HL128550. He is supported also in part by the Evans Medical Foundation and the Jay and Louis Coffman Endowment from the Department of Medicine, Boston University School of Medicine.

Disclosures: Dr. McNally consults for Amgen, AstraZeneca, Avidity, 4D Molecular Therapeutics, Cytokinetics, Janssen, Pfizer, Tenaya, Invitae Corp and Exonics. She is the

founder of Ikaika Therapeutics. These activities are unrelated to the content of this manuscript. Dr. Thanassoulis has participated in advisory boards and speaker bureaus for Amgen, Regeneron/Sanofi, HLS therapeutics, Novartis and Silence therapeutics.

Supplemental Material: Additional Information on Algorithmic and Computational Acceleration of Standard Genetic Models for Imaging Genomics Analyses

REFERENCES

1. Yang Q, Zhong Y, Gillespie C, Merritt R, Bowman B, George MG and Flanders WD. Assessing potential population impact of statin treatment for primary prevention of atherosclerotic cardiovascular diseases in the USA: population-based modelling study. *BMJ Open*. 2017;7:e011684.
2. Wang TW, Asman K, Gentzke AS, Cullen KA, Holder-Hayes E, Reyes-Guzman C, Jamal A, Neff L and King BA. Tobacco Product Use Among Adults - United States, 2017. *MMWR Morb Mortal Wkly Rep*. 2018;67:1225-1232.
3. Hingorani AD and Psaty BM. Primary prevention of cardiovascular disease: time to get more or less personal? *JAMA*. 2009;302:2144-5.
4. Tajeu GS, Booth JN, 3rd, Colantonio LD, Gottesman RF, Howard G, Lackland DT, O'Brien EC, Oparil S, Ravenell J, Safford MM, et al. Incident Cardiovascular Disease Among Adults With Blood Pressure <140/90 mm Hg. *Circulation*. 2017;136:798-812.
5. Musunuru K and Kathiresan S. Genetics of Common, Complex Coronary Artery Disease. *Cell*. 2019;177:132-145.
6. Shen L and Thompson PM. Brain Imaging Genomics: Integrated Analysis and Machine Learning. *Proc IEEE Inst Electr Electron Eng*. 2020;108:125-162.
7. National Heart, Lung, and Blood Institute. Harnessing the New Frontier of Imaging Genomics Workshop for Heart, Lung, Blood and Sleep Disorders. https://www.nhlbi.nih.gov/events/2019/harnessing-new-frontier-imaging-genomics-workshop-heart-lung-blood-and-sleep-disorders. Published 2019. Accessed February 2, 2021.
8. Nicol ED, Norgaard BL, Blanke P, Ahmadi A, Weir-McCall J, Horvat PM, Han K, Bax JJ and Leipsic J. The Future of Cardiovascular Computed Tomography: Advanced Analytics and Clinical Insights. *JACC Cardiovasc Imaging*. 2019;12:1058-1072.
9. Song Y, Wu H, Wen D, Zhu B, Graner P, Ciancibello L, Rajeswaran H, Salem K, Hajmomenian M, Gilkeson RC, et al. Detection of coronary calcifications with dual energy chest X-rays: clinical evaluation. *Int J Cardiovasc Imaging*. 2021;37:767-774.
10. Greenland P, Blaha MJ, Budoff MJ, Erbel R and Watson KE. Coronary Calcium Score and Cardiovascular Risk. *J Am Coll Cardiol*. 2018;72:434-447.
11. Thanassoulis G, Campbell CY, Owens DS, Smith JG, Smith AV, Peloso GM, Kerr KF, Pechlivanis S, Budoff MJ, Harris TB, et al. Genetic associations with valvular calcification and aortic stenosis. *N Engl J Med*. 2013;368:503-12.
12. O'Donnell CJ, Kavousi M, Smith AV, Kardia SL, Feitosa MF, Hwang SJ, Sun YV, Province MA, Aspelund T, Dehghan A, et al. Genome-wide association study for coronary artery calcification with follow-up in myocardial infarction. *Circulation*. 2011;124:2855-64.
13. Natarajan P, Bis JC, Bielak LF, Cox AJ, Dorr M, Feitosa MF, Franceschini N, Guo X, Hwang SJ, Isaacs A, et al. Multiethnic Exome-Wide Association Study of Subclinical Atherosclerosis. *Circ Cardiovasc Genet*. 2016;9:511-520.
14. Malhotra R, Mauer AC, Lino Cardenas CL, Guo X, Yao J, Zhang X, Wunderer F, Smith AV, Wong Q, Pechlivanis S, et al. HDAC9 is implicated in atherosclerotic aortic calcification and affects vascular smooth muscle cell phenotype. *Nat Genet*. 2019;51:1580-1587.
15. Cairns BJ, Coffey S, Travis RC, Prendergast B, Green J, Engert JC, Lathrop M, Thanassoulis G and Clarke R. A Replicated, Genome-Wide Significant Association of Aortic Stenosis With a Genetic Variant for Lipoprotein(a): Meta-Analysis of Published and Novel Data. *Circulation*. 2017;135:1181-1183.
16. Capoulade R, Chan KL, Yeang C, Mathieu P, Bosse Y, Dumesnil JG, Tam JW, Teo KK, Mahmut A, Yang X, et al. Oxidized Phospholipids, Lipoprotein(a), and Progression of Calcific Aortic Valve Stenosis. *J Am Coll Cardiol*. 2015;66:1236-1246.
17. Dweck MR, Maurovich-Horvat P, Leiner T, Cosyns B, Fayad ZA, Gijsen FJH, Van der Heiden K, Kooi ME, Maehara A, Muller JE, et al. Contemporary rationale for non-invasive

imaging of adverse coronary plaque features to identify the vulnerable patient: a Position Paper from the European Society of Cardiology Working Group on Atherosclerosis and Vascular Biology and the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging*. 2020;21:1177-1183.
18. SCOT-HEART Investigators, Newby DE, Adamson PD, Berry C, Boon NA, Dweck MR, Flather M, Forbes J, Hunter A, Lewis S, et al. Coronary CT Angiography and 5-Year Risk of Myocardial Infarction. *N Engl J Med*. 2018;379:924-933.
19. Suh B, Shin DW, Lee SP, Lee H, Lee H, Park EA and Cho B. Family history of coronary heart disease is more strongly associated with coronary than with carotid atherosclerosis in healthy asymptomatic adults. *Atherosclerosis*. 2014;233:584-589.
20. Christiansen MK, Jensen JM, Norgaard BL, Dey D, Botker HE and Jensen HK. Coronary Plaque Burden and Adverse Plaque Characteristics Are Increased in Healthy Relatives of Patients With Early Onset Coronary Artery Disease. *JACC Cardiovasc Imaging*. 2017;10:1128-1135.
21. Villard E, Perret C, Gary F, Proust C, Dilanian G, Hengstenberg C, Ruppert V, Arbustini E, Wichter T, Germain M, et al. A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. *Eur Heart J.* 2011;32:1065-76.
22. Wild PS, Felix JF, Schillert A, Teumer A, Chen M-H, Leening MJG, Völker U, Großmann V, Brody JA, Irvin MR, et al. Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function. *J Clin Invest*. 2017;127:1798-1812.
23. Fox ER, Musani SK, Barbalic M, Lin H, Yu B, Ogunyankin KO, Smith NL, Kutlar A, Glazer NL, Post WS, et al. Genome-Wide Association Study of Cardiac Structure and Systolic Function in African Americans. *Circ Cardiovasc Genet*. 2013;6:37-46.
24. Vasan RS, Glazer NL, Felix JF, Lieb W, Wild PS, Felix SB, Watzinger N, Larson MG, Smith NL, Dehghan A, et al. Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *JAMA*. 2009;302:168-78.
25. Mosley JD, Levinson RT, Farber-Eger E, Edwards TL, Hellwege JN, Hung AM, Giri A, Shuey MM, Shaffer CM, Shi M, et al. The polygenic architecture of left ventricular mass mirrors the clinical epidemiology. *Sci Rep*. 2020;10:7561.
26. Herman DS, Lam L, Taylor MR, Wang L, Teekakirikul P, Christodoulou D, Conner L, DePalma SR, McDonough B, Sparks E, et al. Truncations of titin causing dilated cardiomyopathy. *N Engl J Med*. 2012;366:619-28.
27. Golbus JR, Puckelwartz MJ, Fahrenbach JP, Dellefave-Castillo LM, Wolfgeher D and McNally EM. Population-based variation in cardiomyopathy genes. *Circ Cardiovasc Genet*. 2012;5:391-9.
28. Pirruccello JP, Bick A, Chaffin M, Aragam KG, Choi SH, Lubitz SA, Ho CY, Ng K, Philippakis A, Ellinor PT, et al. Titin Truncating Variants in Adults Without Known Congestive Heart Failure. *J Am Coll Cardiol*. 2020;75:1239-1241.
29. Haggerty CM, Damrauer SM, Levin MG, Birtwell D, Carey DJ, Golden AM, Hartzel DN, Hu Y, Judy R, Kelly MA, et al. Genomics-First Evaluation of Heart Disease Associated With Titin-Truncating Variants. *Circulation*. 2019;140:42-54.
30. McNally EM and Mestroni L. Dilated Cardiomyopathy: Genetic Determinants and Mechanisms. *Circ Res*. 2017;121:731-748.
31. Amzulescu MS, De Craene M, Langet H, Pasquet A, Vancraeynest D, Pouleur AC, Vanoverschelde JL and Gerber BL. Myocardial strain imaging: review of general principles, validation, and sources of discrepancies. *Eur Heart J Cardiovasc Imaging*. 2019;20:605-619.
32. Hiremath P, Bauer M, Cheng HW, Unno K, Liao R and Cheng S. Ultrasonic assessment of myocardial microstructure. *J Vis Exp*. 2014:e50850.
33. Bluemke DA, Kronmal RA, Lima JA, Liu K, Olson J, Burke GL and Folsom AR. The Relationship of Left Ventricular Mass and Geometry to Incident Cardiovascular Events: The MESA (Multi-Ethnic Study of Atherosclerosis) Study. *J Amer Coll Cardiol*. 2008;52:2148-2155.

34. Bai W, Suzuki H, Huang J, Francis C, Wang S, Tarroni G, Guitton F, Aung N, Fung K, Petersen SE, et al. A population-based phenome-wide association study of cardiac and aortic structure and function. *Nat Med*. 2020;26:1654-1662.
35. Pirruccello JP, Bick A, Wang M, Chaffin M, Friedman S, Yao J, Guo X, Venkatesh BA, Taylor KD, Post WS, et al. Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. *Nat Commun*. 2020;11:2254.
36. Ford TJ, Corcoran D, Padmanabhan S, Aman A, Rocchiccioli P, Good R, McEntegart M, Maguire JJ, Watkins S, Eteiba H, et al. Genetic dysregulation of endothelin-1 is implicated in coronary microvascular dysfunction. *Eur Heart J*. 2020;41:3239-3252.
37. Georgiopoulos G, Figliozzi S, Sanguineti F, Aquaro GD, di Bella G, Stamatelopoulos K, Chiribiri A, Garot J, Masci PG and Ismail TF. Prognostic Impact of Late Gadolinium Enhancement by Cardiovascular Magnetic Resonance in Myocarditis: A Systematic Review and Meta-Analysis. *Circ Cardiovasc Imaging*. 2021;14:e011492.
38. Wang TKM, Brizneda MV, Kwon DH, Popovic ZB, Flamm SD, Hanna M, Griffin BP and Xu B. Reference Ranges, Diagnostic and Prognostic Utility of Native T1 Mapping and Extracellular Volume for Cardiac Amyloidosis: A Meta-Analysis. *J Magn Reson Imaging*. 2021;53:1458-1468.
39. Divakaran S, Stewart GC, Lakdawala NK, Padera RF, Zhou W, Desai AS, Givertz MM, Mehra MR, Kwong RY, Hedgire SS, et al. Diagnostic Accuracy of Advanced Imaging in Cardiac Sarcoidosis. *Circ Cardiovasc Imaging.* 2019;12:e008975.
40. Dweck MR, Aikawa E, Newby DE, Tarkin JM, Rudd JH, Narula J and Fayad ZA. Noninvasive Molecular Imaging of Disease Activity in Atherosclerosis. *Circ Res*. 2016;119:330-40.
41. Zheng KH, Kaiser Y, van Olden CC, Santos RD, Dasseux JL, Genest J, Gaudet D, Westerink J, Keyserling C, Verberne HJ, et al. No benefit of HDL mimetic CER-001 on carotid atherosclerosis in patients with genetically determined very low HDL levels. *Atherosclerosis*. 2020;311:13-19.
42. Alam SR, Shah AS, Richards J, Lang NN, Barnes G, Joshi N, MacGillivray T, McKillop G, Mirsadraee S, Payne J, et al. Ultrasmall superparamagnetic particles of iron oxide in patients with acute myocardial infarction: early clinical experience. *Circ Cardiovasc Imaging.* 2012;5:559-65.
43. Joshi NV, Vesey AT, Williams MC, Shah ASV, Calvert PA, Craighead FHM, Yeoh SE, Wallace W, Salter D, Fletcher AM, et al. 18F-fluoride positron emission tomography for identification of ruptured and high-risk coronary atherosclerotic plaques: a prospective clinical trial. *Lancet*. 2014;383:705-713.
44. Tuunanen H, Kuusisto J, Toikka J, Jaaskelainen P, Marjamaki P, Peuhkurinen K, Viljanen T, Sipola P, Stolen KQ, Hannukainen J, et al. Myocardial perfusion, oxidative metabolism, and free fatty acid uptake in patients with hypertrophic cardiomyopathy attributable to the Asp175Asn mutation in the alpha-tropomyosin gene: a positron emission tomography study. *J Nucl Cardiol*. 2007;14:354-65.
45. Parbhudayal RY, Harms HJ, Michels M, van Rossum AC, Germans T and van der Velden J. Increased Myocardial Oxygen Consumption Precedes Contractile Dysfunction in Hypertrophic Cardiomyopathy Caused by Pathogenic TNNT2 Gene Variants. *J Am Heart Assoc*. 2020;9:e015316.
46. Zelt JGE, deKemp RA, Rotstein BH, Nair GM, Narula J, Ahmadi A, Beanlands RS and Mielniczuk LM. Nuclear Imaging of the Cardiac Sympathetic Nervous System: A Disease-Specific Interpretation in Heart Failure. *JACC Cardiovasc Imaging*. 2020;13:1036-1054.
47. Washko GR, Hunninghake GM, Fernandez IE, Nishino M, Okajima Y, Yamashiro T, Ross JC, Estepar RS, Lynch DA, Brehm JM, et al. Lung volumes and emphysema in smokers with interstitial lung abnormalities. *N Engl J Med*. 2011;364:897-906.

48. Oelsner EC, Ortega VE, Smith BM, Nguyen JN, Manichaikul AW, Hoffman EA, Guo X, Taylor KD, Woodruff PG, Couper DJ, et al. A Genetic Risk Score Associated with Chronic Obstructive Pulmonary Disease Susceptibility and Lung Structure on Computed Tomography. *Am J Respir Crit Care Med*. 2019;200:721-731.
49. Wang M, Aaron CP, Madrigano J, Hoffman EA, Angelini E, Yang J, Laine A, Vetterli TM, Kinney PL, Sampson PD, et al. Association Between Long-term Exposure to Ambient Air Pollution and Change in Quantitatively Assessed Emphysema and Lung Function. *JAMA*. 2019;322:546-556.
50. DeMeo DL, Hersh CP, Hoffman EA, Litonjua AA, Lazarus R, Sparrow D, Benditt JO, Criner G, Make B, Martinez FJ, et al. Genetic determinants of emphysema distribution in the national emphysema treatment trial. *Am J Respir Crit Care Med*. 2007;176:42-8.
51. Manichaikul A, Hoffman EA, Smolonska J, Gao W, Cho MH, Baumhauer H, Budoff M, Austin JH, Washko GR, Carr JJ, et al. Genome-wide study of percent emphysema on computed tomography in the general population. The Multi-Ethnic Study of Atherosclerosis Lung/SNP Health Association Resource Study. *Am J Respir Crit Care Med*. 2014;189:408-18.
52. Kong X, Cho MH, Anderson W, Coxson HO, Muller N, Washko G, Hoffman EA, Bakke P, Gulsvik A, Lomas DA, et al. Genome-wide association study identifies BICD1 as a susceptibility gene for emphysema. *Am J Respir Crit Care Med*. 2011;183:43-9.
53. Parker MM, Hao Y, Guo F, Pham B, Chase R, Platig J, Cho MH, Hersh CP, Thannickal VJ, Crapo J, et al. Identification of an emphysema-associated genetic variant near TGFB2 with regulatory effects in lung fibroblasts. *Elife*. 2019;8:e42720.
54. Castaldi PJ, Cho MH, San Jose Estepar R, McDonald ML, Laird N, Beaty TH, Washko G, Crapo JD, Silverman EK and Investigators CO. Genome-wide association identifies regulatory Loci associated with distinct local histogram emphysema patterns. *Am J Respir Crit Care Med*. 2014;190:399-409.
55. Boueiz A, Lutz SM, Cho MH, Hersh CP, Bowler RP, Washko GR, Halper-Stromberg E, Bakke P, Gulsvik A, Laird NM, et al. Genome-Wide Association Study of the Genetic Determinants of Emphysema Distribution. *Am J Respir Crit Care Med*. 2017;195:757-771.
56. Dijkstra AE, Postma DS, van Ginneken B, Wielputz MO, Schmidt M, Becker N, Owsijewitsch M, Kauczor HU, de Koning HJ, Lammers JW, et al. Novel genes for airway wall thickness identified with combined genome-wide association and expression analyses. *Am J Respir Crit Care Med*. 2015;191:547-56.
57. Cho MH, Castaldi PJ, Hersh CP, Hobbs BD, Barr RG, Tal-Singer R, Bakke P, Gulsvik A, San Jose Estepar R, Van Beek EJ, et al. A Genome-Wide Association Study of Emphysema and Airway Quantitative Imaging Phenotypes. *Am J Respir Crit Care Med*. 2015;192:559-69.
58. Sakornsakolpat P, Prokopenko D, Lamontagne M, Reeve NF, Guyatt AL, Jackson VE, Shrine N, Qiao D, Bartz TM, Kim DK, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet*. 2019;51:494-505.
59. Hatabu H, Hunninghake GM, Richeldi L, Brown KK, Wells AU, Remy-Jardin M, Verschakelen J, Nicholson AG, Beasley MB, Christiani DC, et al. Interstitial lung abnormalities detected incidentally on CT: a Position Paper from the Fleischner Society. *Lancet Respir Med*. 2020;8:726-737.
60. Vancheri C, Failla M, Crimi N and Raghu G. Idiopathic pulmonary fibrosis: a disease with similarities and links to cancer biology. *Eur Respir J*. 2010;35:496-504.
61. Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE, Fingerlin TE, Zhang W, Gudmundsson G, Groshong SD, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med*. 2011;364:1503-12.
62. Hunninghake GM, Hatabu H, Okajima Y, Gao W, Dupuis J, Latourelle JC, Nishino M, Araki T, Zazueta OE, Kurugol S, et al. MUC5B promoter polymorphism and interstitial lung abnormalities. *N Engl J Med*. 2013;368:2192-200.

63. Putman RK, Gudmundsson G, Axelsson GT, Hida T, Honda O, Araki T, Yanagawa M, Nishino M, Miller ER, Eiriksdottir G, et al. Imaging Patterns are Associated with Interstitial Lung Abnormality Progression and Mortality. *Am J Respir Crit Care Med*. 2019;200:175-183.
64. Hobbs BD, Putman RK, Araki T, Nishino M, Gudmundsson G, Gudnason V, Eiriksdottir G, Zilhao Nogueira NR, Dupuis J, Xu H, et al. Overlap of Genetic Risk Between Interstitial Lung Abnormalities and Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. 2019;200:1402-1413.
65. Mathai SK, Humphries S, Kropski JA, Blackwell TS, Powers J, Walts AD, Markin C, Woodward J, Chung JH, Brown KK, et al. MUC5B variant is associated with visually and quantitatively detected preclinical pulmonary fibrosis. *Thorax*. 2019;74:1131-1139.
66. Salisbury ML, Hewlett JC, Ding G, Markin CR, Douglas K, Mason W, Guttentag A, Phillips JA, 3rd, Cogan JD, Reiss S, et al. Development and Progression of Radiologic Abnormalities in Individuals at Risk for Familial ILD. *Am J Respir Crit Care Med*. 2020;201:1230-1239.
67. Ash SY, Harmouche R, Ross JC, Diaz AA, Hunninghake GM, Putman RK, Onieva J, Martinez FJ, Choi AM, Lynch DA, et al. The Objective Identification and Quantification of Interstitial Lung Abnormalities in Smokers. *Acad Radiol*. 2017;24:941-946.
68. Desogere P, Tapias LF, Hariri LP, Rotile NJ, Rietz TA, Probst CK, Blasi F, Day H, Mino-Kenudson M, Weinreb P, et al. Type I collagen-targeted PET probe for pulmonary fibrosis detection and staging in preclinical models. *Sci Transl Med*. 2017;9:eaaf4696.
69. Chen HH, Waghorn PA, Wei L, Tapias LF, Schuhle DT, Rotile NJ, Jones CM, Looby RJ, Zhao G, Elliott JM, et al. Molecular imaging of oxidized collagen quantifies pulmonary and hepatic fibrogenesis. *JCI Insight*. 2017;2:e91506.
70. Montesi SB, Izquierdo-Garcia D, Désogère P, Abston E, Liang LL, Digumarthy S, Seethamraju R, Lanuti M, Caravan P and Catana C. Type I Collagen-targeted Positron Emission Tomography Imaging in Idiopathic Pulmonary Fibrosis: First-in-Human Studies. *Am J Respir Crit Care Med*. 2019;200:258-261.
71. Horan GS, Wood S, Ona V, Li DJ, Lukashev ME, Weinreb PH, Simon KJ, Hahm K, Allaire NE, Rinaldi NJ, et al. Partial inhibition of integrin alpha(v)beta6 prevents pulmonary fibrosis without exacerbating inflammation. *Am J Respir Crit Care Med*. 2008;177:56-65.
72. Derlin T, Jaeger B, Jonigk D, Apel R, Freise J, Shin HO, Weiberg D, Warnecke G, Ross TL, Wester HJ, et al. Clinical Molecular Imaging of Pulmonary CXCR4 Expression to Predict Outcome of Pirfenidone Treatment in IPF. *Chest*. 2020;159:1094-1106.
73. Borie R, Fabre A, Prost F, Marchal-Somme J, Lebtahi R, Marchand-Adam S, Aubier M, Soler P and Crestani B. Activation of somatostatin receptors attenuates pulmonary fibrosis. *Thorax*. 2008;63:251-258.
74. Lebtahi R, Moreau S, Marchand-Adam S, Debray MP, Brauner M, Soler P, Marchal J, Raguin O, Gruaz-Guyon A, Reubi JC, et al. Increased uptake of 111In-octreotide in idiopathic pulmonary fibrosis. *J Nucl Med*. 2006;47:1281-1287.
75. Ambrosini V, Zompatori M, De Luca F, Antonia D, Allegri V, Nanni C, Malvi D, Tonveronachi E, Fasano L, Fabbri M, et al. 68Ga-DOTANOC PET/CT allows somatostatin receptor imaging in idiopathic pulmonary fibrosis: preliminary results. *J Nucl Med*. 2010;51:1950-5.
76. Montesi SB, Rao R, Liang LL, Goulart HE, Sharma A, Digumarthy SR, Shea BS, Seethamraju RT, Caravan P and Tager AM. Gadofosveset-enhanced lung magnetic resonance imaging to detect ongoing vascular leak in pulmonary fibrosis. *Eur Respir J*. 2018;51:1800171.
77. Shea BS, Brooks SF, Fontaine BA, Chun J, Luster AD and Tager AM. Prolonged exposure to sphingosine 1-phosphate receptor-1 agonists exacerbates vascular leak, fibrosis, and mortality after lung injury. *Am J Respir Cell Mol Biol*. 2010;43:662-73.
78. Win T, Screaton NJ, Porter JC, Ganeshan B, Maher TM, Fraioli F, Endozo R, Shortman RI, Hurrell L, Holman BF, et al. Pulmonary (18)F-FDG uptake helps refine current risk

stratification in idiopathic pulmonary fibrosis (IPF). *Eur J Nucl Med Mol Imaging*. 2018;45:806-815.
79. Bondue B, Castiaux A, Van Simaeys G, Mathey C, Sherer F, Egrise D, Lacroix S, Huaux F, Doumont G and Goldman S. Absence of early metabolic response assessed by 18F-FDG PET/CT after initiation of antifibrotic drugs in IPF patients. *Respir Res*. 2019;20:10.
80. Withana NP, Ma X, McGuire HM, Verdoes M, van der Linden WA, Ofori LO, Zhang R, Li H, Sanman LE, Wei K, et al. Non-invasive Imaging of Idiopathic Pulmonary Fibrosis Using Cathepsin Protease Probes. *Sci Rep*. 2016;6:19755.
81. Russell S and Norvig P. *Artificial Intelligence: A Modern Approach*: Pearson; 2016.
82. Mathur P, Srivastava S, Xu X and Mehta JL. Artificial Intelligence, Machine Learning, and Cardiovascular Disease. *Clin Med Insights Cardiol*. 2020;14:1179546820927404.
83. Kolossvary M, De Cecco CN, Feuchtner G and Maurovich-Horvat P. Advanced atherosclerosis imaging by CT: Radiomics, machine learning and deep learning. *J Cardiovasc Comput Tomogr*. 2019;13:274-280.
84. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, Heidenreich PA, Harrington RA, Liang DH, Ashley EA, et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature*. 2020;580:252-256.
85. Fries JA, Varma P, Chen VS, Xiao K, Tejeda H, Saha P, Dunnmon J, Chubb H, Maskatia S, Fiterau M, et al. Weakly supervised classification of aortic valve malformations using unlabeled cardiac MRI sequences. *Nat Commun*. 2019;10:3111.
86. Meyer HV, Dawes TJW, Serrani M, Bai W, Tokarczuk P, Cai J, de Marvao A, Henry A, Lumbers RT, Gierten J, et al. Genetic and functional insights into the fractal structure of the heart. *Nature*. 2020;584:589-594.
87. Cordova-Palomera A, Tcheandjieu C, Fries JA, Varma P, Chen VS, Fiterau M, Xiao K, Tejeda H, Keavney BD, Cordell HJ, et al. Cardiac Imaging of Aortic Valve Area From 34 287 UK Biobank Participants Reveals Novel Genetic Associations and Shared Genetic Comorbidity With Multiple Disease Phenotypes. *Circ Genom Precis Med*. 2020;13:e003014.
88. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, Hamzah H, Garcia-Franco R, San Yeo IY, Lee SY, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA*. 2017;318:2211-2223.
89. Meyer M, Ronald J, Vernuccio F, Nelson RC, Ramirez-Giraldo JC, Solomon J, Patel BN, Samei E and Marin D. Reproducibility of CT Radiomic Features within the Same Patient: Influence of Radiation Dose and CT Reconstruction Settings. *Radiology*. 2019;293:583-591.
90. Berenguer R, Pastor-Juan MDR, Canales-Vazquez J, Castro-Garcia M, Villas MV, Mansilla Legorburo F and Sabater S. Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. *Radiology*. 2018;288:407-415.
91. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ and Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med*. 2018;15:e1002683.
92. Noseworthy PA, Attia ZI, Brewer LC, Hayes SN, Yao X, Kapa S, Friedman PA and Lopez-Jimenez F. Assessing and Mitigating Bias in Medical Artificial Intelligence: The Effects of Race and Ethnicity on a Deep Learning Model for ECG Analysis. *Circ Arrhythm Electrophysiol*. 2020;13:e007988.
93. Lu MT, Raghu VK, Mayrhofer T, Aerts H and Hoffmann U. Deep Learning Using Chest Radiographs to Identify High-Risk Smokers for Lung Cancer Screening Computed Tomography: Development and Validation of a Prediction Model. *Ann Intern Med*. 2020;173:704-713.
94. Lu MT, Ivanov A, Mayrhofer T, Hosny A, Aerts H and Hoffmann U. Deep Learning to Assess Long-term Mortality From Chest Radiographs. *JAMA Netw Open*. 2019;2:e197416.

95. Kolossvary M, Kellermayer M, Merkely B and Maurovich-Horvat P. Cardiac Computed Tomography Radiomics: A Comprehensive Review on Radiomic Techniques. *J Thorac Imaging*. 2018;33:26-34.
96. Kolossvary M, Karady J, Szilveszter B, Kitslaar P, Hoffmann U, Merkely B and Maurovich-Horvat P. Radiomic Features Are Superior to Conventional Quantitative Computed Tomographic Metrics to Identify Coronary Plaques With Napkin-Ring Sign. *Circ Cardiovasc Imaging*. 2017;10:e006843.
97. Kolossvary M, Karady J, Kikuchi Y, Ivanov A, Schlett CL, Lu MT, Foldyna B, Merkely B, Aerts HJ, Hoffmann U, et al. Radiomics versus Visual and Histogram-based Assessment to Identify Atheromatous Lesions at Coronary CT Angiography: An ex Vivo Study. *Radiology*. 2019;293:89-96.
98. Baessler B, Mannil M, Oebel S, Maintz D, Alkadhi H and Manka R. Subacute and Chronic Left Ventricular Myocardial Scar: Accuracy of Texture Analysis on Nonenhanced Cine MR Images. *Radiology*. 2018;286:103-112.
99. Neisius U, El-Rewaidy H, Nakamori S, Rodriguez J, Manning WJ and Nezafat R. Radiomic Analysis of Myocardial Native T1 Imaging Discriminates Between Hypertensive Heart Disease and Hypertrophic Cardiomyopathy. *JACC Cardiovasc Imaging*. 2019;12:1946-1954.
100. Yang J, Angelini ED, Balte PP, Hoffman EA, Austin JH, Smith BM, Barr RG and Laine AF. Novel Subtypes of Pulmonary Emphysema Based on Spatially-Informed Lung Texture Learning. *arXiv preprint arXiv:200704978*. 2020.
101. Humphries SM, Notary AM, Centeno JP, Strand MJ, Crapo JD, Silverman EK, Lynch DA and Genetic Epidemiology of CI. Deep Learning Enables Automatic Classification of Emphysema Pattern at CT. *Radiology*. 2020;294:434-444.
102. Young AL, Bragman FJS, Rangelov B, Han MK, Galban CJ, Lynch DA, Hawkes DJ, Alexander DC, Hurst JR and Investigators CO. Disease Progression Modeling in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 2020;201:294-302.
103. Pickhardt PJ, Graffy PM, Zea R, Lee SJ, Liu J, Sandfort V and Summers RM. Automated CT biomarkers for opportunistic prediction of future cardiovascular events and mortality in an asymptomatic screening population: a retrospective cohort study. *Lancet Digit Health*. 2020;2:e192-e200.
104. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A, Gogarten SM, Kang HM, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021;590:290-299.
105. Asimit JL, Hatzikotoulas K, McCarthy M, Morris AP and Zeggini E. Trans-ethnic study design approaches for fine-mapping. *Eur J Hum Genet*. 2016;24:1330-6.
106. Cresci S, Pereira NL, Ahmad F, Byku M, de Las Fuentes L, Lanfear DE, Reilly CM, Owens AT and Wolf MJ. Heart Failure in the Era of Precision Medicine: A Scientific Statement From the American Heart Association. *Circ Genom Precis Med*. 2019;12:458-485.
107. Andersson C, Lin H, Liu C, Levy D, Mitchell GF, Larson MG and Vasan RS. Integrated Multiomics Approach to Identify Genetic Underpinnings of Heart Failure and Its Echocardiographic Precursors: Framingham Heart Study. *Circ Genom Precis Med*. 2019;12:e002489.
108. Ibrahim NE and Januzzi JL, Jr. Established and Emerging Roles of Biomarkers in Heart Failure. *Circ Res*. 2018;123:614-629.
109. Smith JG and Gerszten RE. Emerging Affinity-Based Proteomic Technologies for Large-Scale Plasma Profiling in Cardiovascular Disease. *Circulation*. 2017;135:1651-1664.
110. Tromp J, Hage C, Ouwerkerk W, Sanders-van Wijk S, Svedlund S, Saraste A, Ljung Faxen U, Lagerstrom Fermer M, Gan LM, Lund LH, et al. Biomarker Correlates of Coronary Microvascular Dysfunction in Heart Failure With Preserved Ejection Fraction. *Circulation*. 2019;140:1359-1361.

111. Sanders-van Wijk S, Tromp J, Beussink-Nelson L, Hage C, Svedlund S, Saraste A, Swat SA, Sanchez C, Njoroge J, Tan RS, et al. Proteomic Evaluation of the Comorbidity-Inflammation Paradigm in Heart Failure with Preserved Ejection Fraction: Results from the PROMIS-HFpEF Study. *Circulation*. 2020;142:2029-2044.
112. Nayor M, Short MI, Rasheed H, Lin H, Jonasson C, Yang Q, Hveem K, Felix JF, Morrison AC, Wild PS, et al. Aptamer-Based Proteomic Platform Identifies Novel Protein Predictors of Incident Heart Failure and Echocardiographic Traits. *Circ Heart Fail*. 2020;13:e006749.
113. Cheng S, Shah SH, Corwin EJ, Fiehn O, Fitzgerald RL, Gerszten RE, Illig T, Rhee EP, Srinivas PR, Wang TJ, et al. Potential Impact and Study Considerations of Metabolomics in Cardiovascular Health and Disease: A Scientific Statement From the American Heart Association. *Circ Cardiovasc Genet*. 2017;10:e000032.
114. Neubauer S. The failing heart--an engine out of fuel. *N Engl J Med*. 2007;356:1140-1151.
115. Ahmad T, Kelly JP, McGarrah RW, Hellkamp AS, Fiuzat M, Testani JM, Wang TS, Verma A, Samsky MD, Donahue MP, et al. Prognostic Implications of Long-Chain Acylcarnitines in Heart Failure and Reversibility With Mechanical Circulatory Support. *J Am Coll Cardiol*. 2016;67:291-9.
116. Elmariah S, Farrell LA, Furman D, Lindman BR, Shi X, Morningstar JE, Rhee EP and Gerszten RE. Association of Acylcarnitines With Left Ventricular Remodeling in Patients With Severe Aortic Stenosis Undergoing Transcatheter Aortic Valve Replacement. *JAMA Cardiol*. 2018;3:242-246.
117. Nwabuo CC, Duncan M, Xanthakis V, Peterson LR, Mitchell GF, McManus D, Cheng S and Vasan RS. Association of Circulating Ceramides With Cardiac Structure and Function in the Community: The Framingham Heart Study. *J Am Heart Assoc*. 2019;8:e013050.
118. Andersson C, Liu C, Cheng S, Wang TJ, Gerszten RE, Larson MG and Vasan RS. Metabolomic signatures of cardiac remodelling and heart failure risk in the community. *ESC Heart Fail*. 2020;7:3707-3715.
119. Rhee EP, Yang Q, Yu B, Liu X, Cheng S, Deik A, Pierce KA, Bullock K, Ho JE, Levy D, et al. An exome array study of the plasma metabolome. *Nat Commun*. 2016;7:12360.
120. Benson MD, Yang Q, Ngo D, Zhu Y, Shen D, Farrell LA, Sinha S, Keyes MJ, Vasan RS, Larson MG, et al. Genetic Architecture of the Cardiovascular Risk Proteome. *Circulation*. 2018;137:1158-1172.
121. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, Burgess S, Jiang T, Paige E, Surendran P, et al. Genomic atlas of the human plasma proteome. *Nature*. 2018;558:73-79.
122. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, Gutteridge A, Erola P, Liu Y, Luo S, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet*. 2020;52:1122-1131.
123. Mosley JD, Benson MD, Smith JG, Melander O, Ngo D, Shaffer CM, Ferguson JF, Herzig MS, McCarty CA, Chute CG, et al. Probing the Virtual Proteome to Identify Novel Disease Biomarkers. *Circulation*. 2018;138:2469-2481.
124. Mahdavi S, Jenkins DJA, Borchers CH and El-Sohemy A. Genetic Variation in 9p21 and the Plasma Proteome. *J Proteome Res*. 2018;17:2649-2656.
125. Ganjgahi H, Wincker A, DC. G, Blangero J, P. K and Nichols T. Fast and Powerful Heritability Inference for Family-Based Neuroimaging Studies. *Neuroimage*. 2015;115:256-68.
126. Ganjgahi H, Winkler AM, Glahn DC, Blangero J, Donohue B, Kochunov P and Nichols TE. Fast and Powerful Genome Wide Association Analysis of Dense Genetic Data with High Dimensional Imaging Phenotypes. *Nat Commun*. 2018;9:3254.

127. Kochunov P, Donohue B, Mitchell BD, Ganjgahi H, Adhikari B, Ryan M, Medland SE, Jahanshad N, Thompson PM, Blangero J, et al. Genomic kinship construction to enhance genetic analyses in the human connectome project data. *Hum Brain Mapp*. 2019;40:1677-1688.
128. Blangero J, Diego VP, Dyer TD, Almeida M, Peralta J, Kent JW, Jr., Williams JT, Almasy L and Goring HH. A kernel of truth: statistical advances in polygenic variance component models for complex human pedigrees. *Adv Genet*. 2013;81:1-31.
129. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*. 2013;9:e1003348.
130. Apostolova LG, Risacher SL, Duran T, Stage EC, Goukasian N, West JD, Do TM, Grotts J, Wilhalme H, Nho K, et al. Associations of the Top 20 Alzheimer Disease Risk Variants With Brain Amyloidosis. *JAMA Neurol*. 2018;75:328-341.
131. Xu Z, Wu C, Pan W and Alzheimer's Disease Neuroimaging I. Imaging-wide association study: Integrating imaging endophenotypes in GWAS. *NeuroImage*. 2017;159:159-169.
132. Yang Q, Wu H, Guo CY and Fox CS. Analyze multivariate phenotypes in genetic association studies by combining univariate association tests. *Genet Epidemiol*. 2010;34:444-54.
133. Lan H, Stoehr JP, Nadler ST, Schueler KL, Yandell BS and Attie AD. Dimension reduction for mapping mRNA abundance as quantitative traits. *Genetics*. 2003;164:1607-14.
134. Ferreira MA and Purcell SM. A multivariate test of association. *Bioinformatics*. 2009;25:132-3.
135. Korte A, Vilhjalmsson BJ, Segura V, Platt A, Long Q and Nordborg M. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genet*. 2012;44:1066-71.
136. Holden M, Deng S, Wojnowski L and Kulle B. GSEA-SNP: applying gene set enrichment analysis to SNP data from genome-wide association studies. *Bioinformatics*. 2008;24:2784-5.
137. Yan J, Risacher SL, Shen L and Saykin AJ. Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. *Brief Bioinform*. 2018;19:1370-1381.
138. Yao X, Yan J, Liu K, Kim S, Nho K, Risacher SL, Greene CS, Moore JH, Saykin AJ, Shen L, et al. Tissue-specific network-based genome wide study of amygdala imaging phenotypes to identify functional interaction modules. *Bioinformatics*. 2017;33:3250-3257.
139. Wang H, Nie F, Huang H, Kim S, Nho K, Risacher SL, Saykin AJ, Shen L and Alzheimer's Disease Neuroimaging I. Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort. *Bioinformatics*. 2012;28:229-37.
140. Vounou M, Nichols TE, Montana G and Alzheimer's Disease Neuroimaging I. Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *NeuroImage*. 2010;53:1147-59.
141. Du L, Jingwen Y, Kim S, Risacher SL, Huang H, Inlow M, Moore JH, Saykin AJ, Shen L and Alzheimer's Disease Neuroimaging I. A novel structure-aware sparse learning algorithm for brain imaging genetics. *Med Image Comput Assist Interv.* 2014;17:329-36.
142. Du L, Huang H, Yan J, Kim S, Risacher SL, Inlow M, Moore JH, Saykin AJ, Shen L and Alzheimer's Disease Neuroimaging I. Structured sparse canonical correlation analysis for brain imaging genetics: an improved GraphNet method. *Bioinformatics*. 2016;32:1544-51.
143. Kauppi K, Fan CC, McEvoy LK, Holland D, Tan CH, Chen CH, Andreassen OA, Desikan RS, Dale AM and Alzheimer's Disease Neuroimaging I. Combining Polygenic Hazard Score With Volumetric MRI and Cognitive Measures Improves Prediction of Progression From Mild Cognitive Impairment to Alzheimer's Disease. *Front Neurosci*. 2018;12:260.
144. Brand L, Nichols K, Wang H, Huang H, Shen L and Alzheimer's Disease Neuroimaging I. Predicting Longitudinal Outcomes of Alzheimer's Disease via a Tensor-Based Joint Classification and Regression Model. *Pac Symp Biocomput*. 2020;25:7-18.

145. Zhou T, Thung KH, Zhu X and Shen D. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Hum Brain Mapp*. 2019;40:1001-1016.
146. Yan J, Risacher SL, Nho K, Saykin AJ and Shen L. Identification of Discriminative Imaging Proteomics Associations in Alzheimer's Disease Via a Novel Sparse Correlation Model. *Pac Symp Biocomput.* 2017;22:94-104.
147. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, Whitbourne S, Deen J, Shannon C, Humphries D, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. 2016;70:214-23.
148. National Institutes of Health. Data Commons. https://commonfund.nih.gov/commons. Published 2019. Accessed February 2, 2021.
149. National Heart, Lung, and Blood Institute. Data STAGE. Storage, Toolspace, Access and analytics for biG data Empowerment. https://www.nhlbi.nih.gov/science/biodata-catalyst. Published 2018. Accessed February 2, 2021.
150. Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, Cornes BK, Montgomery GW and Martin NG. Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings. *PLoS Genetics*. 2006;2:e41.
151. Visscher PeterÂ M, Macgregor S, Benyamin B, Zhu G, Gordon S, Medland S, Hill WilliamÂ G, Hottenga J-J, Willemsen G, Boomsma DorretÂ I, et al. Genome Partitioning of Genetic Variation for Height from 11,214 Sibling Pairs. *Am J Hum Genet.* 2007;81:1104-1110.
152. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42:565-9.
153. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*. 2014;46:1173-86.
154. Toro R, Poline JB, Huguet G, Loth E, Frouin V, Banaschewski T, Barker GJ, Bokde A, Buchel C, Carvalho FM, et al. Genomic architecture of human neuroanatomical diversity. *Mol Psychiatry*. 2014.
155. Ramstetter MD, Dyer TD, Lehman DM, Curran JE, Duggirala R, Blangero J, Mezey JG and Williams AL. Benchmarking Relatedness Inference Methods with Genome-Wide Data from Thousands of Relatives. *Genetics*. 2017;207:75-82.
156. Blangero J, Williams JT and Almasy L. Variance component methods for detecting complex trait loci. *Adv Genet*. 2001;42:151-81.
157. Almasy L and Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet*. 1998;62:1198-211.
158. Kochunov P, Patel B, Ganjgahi H, Donohue B, Ryan M, Hong EL, Chen X, Adhikari B, Jahanshad N, Thompson PM, et al. Homogenizing Estimates of Heritability Among SOLAR-Eclipse, OpenMx, APACE, and FPHI Software Packages in Neuroimaging Data. *Front Neuroinform*. 2019;13:16.
159. Lee D, Dinov I, Dong B, Gutman B, Yanovsky I and Toga AW. CUDA optimization strategies for compute- and memory-bound neuroimaging algorithms. *Comput Methods Programs Biomed*. 2012;106:175-87.

FIGURE LEGENDS

Figure 1. Imaging genomics: learning problems, applications, and statistical and machine learning considerations. Adapted from reference[6]. The key learning problems (or objectives) of imaging genomics are displayed in panel (a). Panel (b) displays potential applications for combining biomedical imaging and state-of-the-art genomics information for disease characterization, diagnosis, and risk prediction. Considerations for machine learning and statistical approaches to imaging genomics analyses are shown in panel (c).
Abbreviations: GWA, genome-wide association; WGS, whole genome sequencing; Dz, disease; surv, survival; dysf, dysfunction; mod, moderate

Figure 2. Novel imaging-based sub-phenotypes of COPD and emphysema. (Reproduced with permission from reference[100].) Spatially-informed lung texture patterns (sLTPs) were derived using unsupervised clustering on chest CT scans from 317 participants in the MESA COPD study based on the texture and spatial location of hundreds of regions-of-interest per scan. Panel (a) shows two examples of lung CTs and colorized maps of their different sLTPs. Panel (b) displays characteristics of the 12 lung texture patterns based on axial cuts from nine random regions of interest (top of each box), the average lung texture pattern calculated from the images (middle of each box), and the spatial density plots (bottom of each box). These lung texture patterns are reproducible and relate to measures of symptom and disease severity.
Abbreviations: sLTP=spatially-informed lung texture patterns; S=superior; I=inferior; P=posterior; A=anterior

Table 1. **Bioimaging modalities for heart and lung diseases**

Imaging modality	Characteristics	Examples of applications in heart and lung diseases
Computed tomography (CT)	-Fast, relatively low radiation dose -Excellent spatial resolution -High throughput, widely available	*Heart:* -Coronary artery and aortic valve calcium (no contrast required) -Coronary artery angiography (with contrast) *Lungs:* -Qualitative disease characterization (i.e., emphysema, ILD) -Quantitative lung measures (e.g., airway wall thickness, interstitial abnormalities, small vessel pruning)
Magnetic resonance imaging (MRI)	-Provides simultaneous assessment of heart structure and function -Evaluation of lung parenchyma with traditional methods is challenging due to low signal intensity -Takes longer than CT, but can be performed efficiently in large cohorts -No radiation required	*Heart:* -Cardiac structure and function (e.g., chamber sizes, ejection fraction) -Fibrosis, iron deposition, presence of scar through specialized sequences *Lung:* -Pulmonary vascular characteristics -Methods in development for lung parenchyma evaluation including different signal sequences and inhalation of gas mixtures serving as contrast agents
Ultrasound (echocardiography)	-Easily obtained and readily available -Radiation free -Excellent characterization of heart structure and function	*Heart:* -Cardiac chamber size measurement -Cardiac function, including quantitative measures such as strain imaging
Molecular imaging	-Precise biological characterization -Can measure disease *activity*, not just organ structure/function -May allow disease identification prior to overt manifestations -Relatively low throughput, challenges with obtaining images at large scale -Higher radiation dose necessary -Largely preclinical	*Heart:* -Atherosclerotic plaque activity with 18F-FDG PET imaging -Microvascular calcification with 18F-fluoride PET tracer -Assessment of myocardial perfusion and metabolism with PET *Lung:* -Different probes can identify steps in development of lung fibrosis including collagen deposition and cross-linking, fibrosis development, fibroblast activity, and inflammation

Abbreviations: ILD, interstitial lung disease; FDG, fluorodeoxyglucose; PET, positron emission tomography

Table 2. Research priorities in imaging genomics

Research Priorities
Methods development in imaging, genomics, and their statistical integration
Infrastructure for data analyses and data sharing
Multiethnic large samples to capture diversity and spectrum of phenotypic and genomic variation of common conditions
Analysis of longitudinal imaging data in relation to longitudinal molecular data
Methods development for testing effect modification and causal inference adapted to imaging genomics
Consortia focused on less common heritable conditions
Defining standards for imaging genomics methods and interpretation norms

Figure 1.

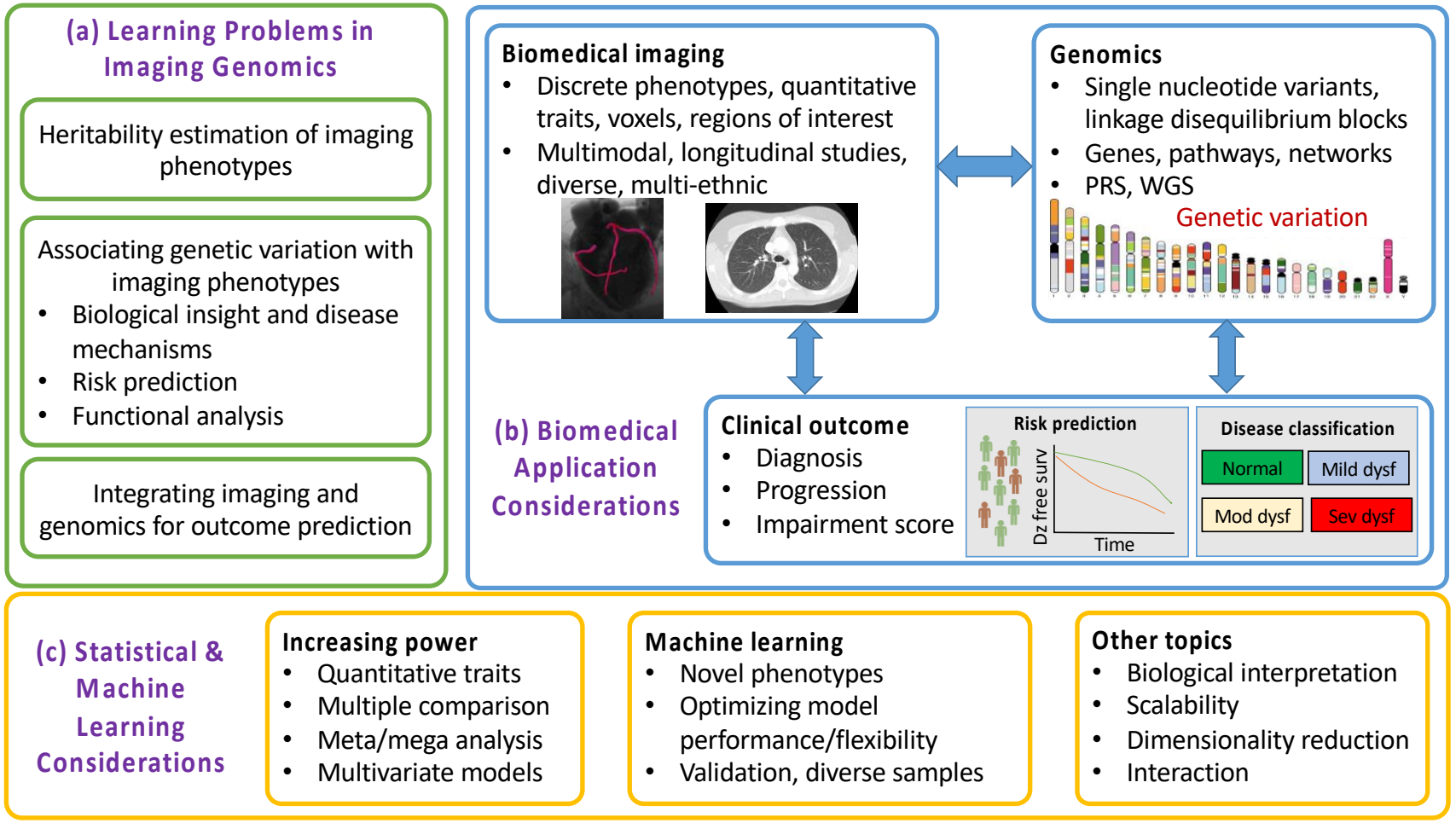
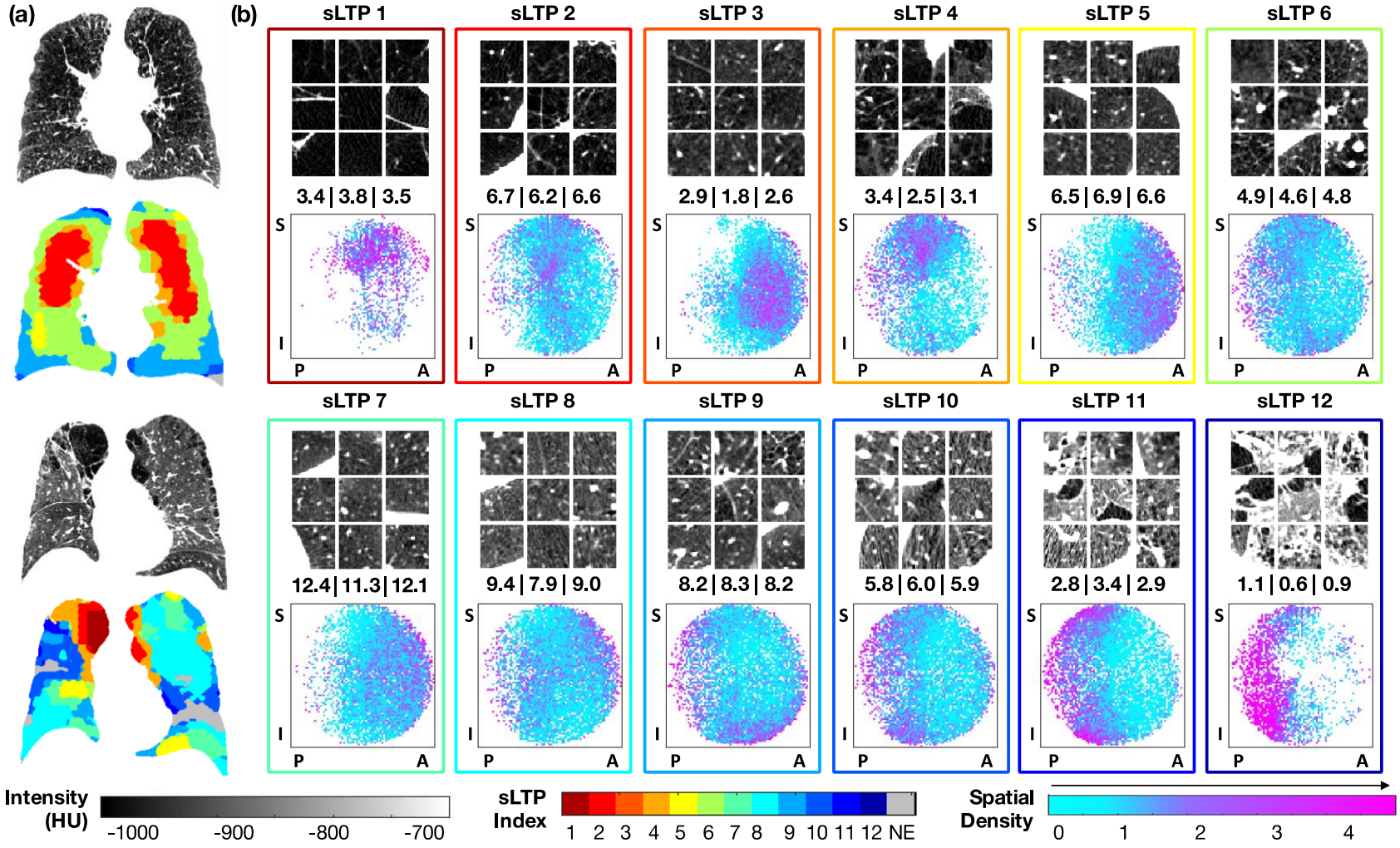

(a) Learning Problems in Imaging Genomics
Heritability estimation of imaging phenotypes
Associating genetic variation with imaging phenotypes
• Biological insight and disease mechanisms
• Risk prediction
• Functional analysis
Integrating imaging and genomics for outcome prediction
Biomedical imaging
• Discrete phenotypes, quantitative traits, voxels, regions of interest
• Multimodal, longitudinal studies, diverse, multi-ethnic
Genomics
• Single nucleotide variants, linkage disequilibrium blocks
• Genes, pathways, networks
• PRS, WGS
Genetic variation
(b) Biomedical Application Considerations
Clinical outcome
• Diagnosis
• Progression
• Impairment score
Risk prediction
Dz free surv
Time
Disease classification
Normal
Mild dysf
Mod dysf
Sev dysf
(c) Statistical & Machine Learning Considerations
Increasing power
• Quantitative traits
• Multiple comparison
• Meta/mega analysis
• Multivariate models
Machine learning
• Novel phenotypes
• Optimizing model performance/flexibility
• Validation, diverse samples
Other topics
• Biological interpretation
• Scalability
• Dimensionality reduction
• Interaction

Figure 2.


(a)
(b)
sLTP 1
sLTP 2
sLTP 3
sLTP 4
sLTP 5
sLTP 6
3.4 | 3.8 | 3.5
6.7 | 6.2 | 6.6
2.9 | 1.8 | 2.6
3.4 | 2.5 | 3.1
6.5 | 6.9 | 6.6
4.9 | 4.6 | 4.8
sLTP 7
sLTP 8
sLTP 9
sLTP 10
sLTP 11
sLTP 12
12.4 | 11.3 | 12.1
9.4 | 7.9 | 9.0
8.2 | 8.3 | 8.2
5.8 | 6.0 | 5.9
2.8 | 3.4 | 2.9
1.1 | 0.6 | 0.9
S
I
P
A
Intensity (HU)
-1000
-900
-800
-700
sLTP Index
1 2 3 4 5 6 7 8 9 10 11 12 NE
Spatial Density
0 1 2 3 4