Specific Aims

Medical research significantly benefits from the development and proliferation of imaging- related analysis packages, particularly those softwares which have been tailored for specific application domains. Although several such established packages exist for the neuroimaging community (e.g., FSL, FreeSurfer, AFNI, SPM), no such package exists for the pulmonary imaging research community. The primary goal of this proposal is to develop and provide a dissemination platform for a robust, open-source image analysis toolkit with annotated data specifically targeted at the pulmonary research community.

Although methodological research is continually being presented at conferences and published in various venues, the unfortunate reality is that much of this work exists strictly in “advertisement” form. Oftentimes the underlying code is unavailable to other researchers or is implemented in a limited manner (i.e., strictly as proof-of-concept software). Frequently crucial parameter choices are omitted in the corresponding publication(s) which makes outside implementations difficult. Additionally, the data used to showcase the proposed methodology are often private and actual data visualization is limited to carefully selected snapshots for publication (i.e., advertisement) purposes which might not be representative of algorithmic performance.

As a potential corrective, this proposal will provide an open-source software toolkit for core pulmonary image analysis tasks across multiple modalities which we proposed in recent publications. These basic tasks include image registration, template building for cross-sectional and longitudinal analyses, and functional and structural lung image segmentation. In addition to the software, we will provide both the input and output data as open-source not only so that other users can it to reproduce our results, but it will also allow researchers to use it in their studies.

As principal developers of the popular, open-source ANTs (Advanced Normalization Tools) package, we have extensive experience in the development of well-written software that has gained much traction in the neuroscience community. We have also participated in several image analysis competitions for a variety of applications (neuro, pulmonary, and cardiac) and data scaling and believe that this will also contribute to our success in accomplishing the goals of this application as defined by the following specific aims:

- **Specific Aim 1.** To develop a set of open-source software tools for CT, proton, and He-3 pulmonary computational analysis.
- **Specific Aim 2.** To provide multiple sets of multi-modal annotated lung data (CT, proton, and He3) for public use.
- ***Specific Aim 3.*** *Perhaps instead of Aim 2, we can include analysis of one of Mike's comprehensive data sets (multi-modal CT, proton & Xe), publish the results, as well as provide all the scripts (using ANTsR?) for the user to re-run the analysis on the user side.*

Say something more here (e.g., We believe that this will be a valuable resource to the community.)

Research Strategy

3(a) Significance. Well-vetted and publicly available software is a significant benefit to various research communities. For example, the neuroscience community has greatly benefited from highly evolved software packages such as FreeSurfer [1], the FMRIB Software Library (FSL) [2], the Analysis of Functional NeuroImages (AFNI) package [3], and the Statistical Parametric Mapping (SPM) package [4]. Performing a pubmed query for any one of these softwares every year for the past decade (cf Figure 1) illustrates the growing use of such packages and the research studies that are produced as a result. However, despite the demonstrable benefits and year-by-year usage increase, no such analogous set of tools exist for pulmonary-specific research.

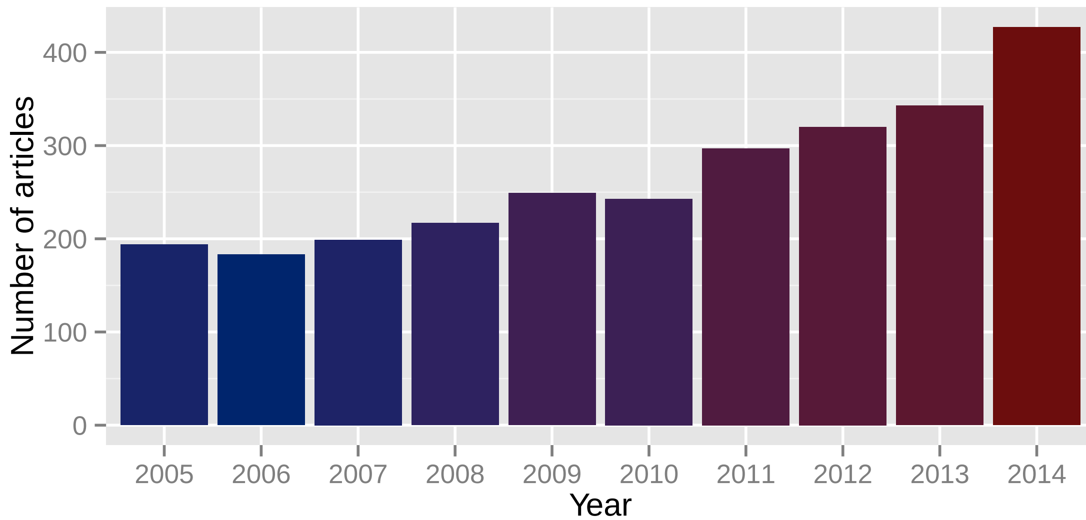

Figure 1: Number of articles per year which use common, publicly available neuroimaging analysis packages.

Medical image analysis libraries (e.g., the Insight ToolKit) provide extensive algorithmic capabilities for a range of generic medical image analysis tasks. However, specific applications (e.g., lung image analysis) are not available despite the vast number of algorithms that have been proposed in the literature. This lack was one of the primary motivations for the development of our Advanced Normalization Tools (ANTs). ANTs takes advantage of the mature Insight ToolKit in providing an optimal software framework for building scripts and programs specifically for neuroimaging. For example, the following core neuroimage processing algorithms have been made available through our ANTs toolkit (complete with examples and developer-tuned parameters) and have been used extensively by our group and others: brain normalization [5, 6], brain template generation [7], skull-stripping or brain extraction [8, 9], prior-based Bayesian brain tissue segmentation [5], cortical gray matter thickness estimation [9], brain tumor segmentation [10], and cortical labeling [11, 12]. In addition to public availability, some of these algorithms have been showcased in international competitions and have performed extremely well [**???**, 13].

Analogously, several algorithmic categories exist for lung image analysis which, as we have stated previously, do not exist in any comprehensive, publicly available package. An extensive survey concentrating on the years 1999–2004 is given in [14] which covers computer aided diagnosis of lung disease and lung cancer in CT (i.e., detection and tracking of pulmonary nodules) and provides an overview of the many relevant segmentation methods for pulmonary structures. Although many algorithms existed at the time, continued technical development has only increased the number of available algorithms. Following is a small sampling of more recently reported techniques for CT analysis:

- whole lung differentiation from the chest wall (e.g., [15–18])
- bronchial structure extraction (e.g.,[19, 20]; the many submissions to the recent Extraction of Airways from CT (ExACT) challenge of the 2nd International Workshop on Pulmonary Image Analysis [21]),
- vasculature segmentation (e.g., [22, 23]),
- lobe and/or fissure detection (e.g., [24, 25]), and
- feature extraction and classification (e.g., [26–28]).

Since this list is restricted to CT image analysis, inclusion of additional techniques specific to other modalities will have additional benefit. For example, ventilation-based segmentation for analysis of ventilation lung imaging [29] will also have significant impact in a comprehensive lung image analysis suite.

Important in any methodological discussion is the crucial importance of parameter selection which requires domain-specific experience. For example, although ANTs performance in brain registration has been independently evaluated and found to be of relatively high quality [Klein2009], applying our registration tool in the EMPIRE10 challenge (Evaluation of Methods for Pulmonary Image REgistration 2010) required significant empirically-based tuning. In addition, new innovations in diffeomorphic registration technology has led to a Symmetric Normalization B-spline variant which has demonstrated preferred normalizations [30], particularly for pulmonary data [31]. Note that the goals of this proposal would significantly support the National Library of Medicine's own open-source directives in that all software would be developed using the established Insight ToolKit's coding and testing standards with the eventual idea that much (if not all) of the actual code would be contributed for inclusion in future versions of the Insight ToolKit.

It should also be noted that open-source software, in general, has documented benefits within the targeted communities for which it is developed and supported. In addition to the increase in research output illustrated earlier, open-source permits students and researchers to learn specific computational techniques in a social environment [32]. This, in turn, provides motivation for user-based support including potential contributions such as bug fixes and feature additions. Additional analyses have shown the tremendous cost savings that open-source software yields [33].

3(b) Innovation. Given the lack of open-source solutions for pulmonary image analysis, the proposal goals would produce an innovative framework for corresponding research. Many algorithms have been proposed in various technical venues but that which we propose would provide well-vetted and easy-to-use implementations of these robust methodologies. Many of these have been developed by our group.

An additional innovation we are proposing is the inclusion of data and detailed instructions for generating a reproducible, multimodality pulmonary study using the proposed package. *Say more about the data specifics here.* Clinical findings will be published in a traditional journal (e.g., Chest) for the interested researcher. In addition, we will provide all image data and the quantitative analysis scripts as a companion release to accompany the paper (e.g., see previous similar offerings from our group [?,?]) Such a comprehensive clinical investigation using these tools will not only provide insight into the specifics of ?? but will also provide a tangible mechanism for using the tools created with this proposal.

3(c) Approach.

Specific Aim 1. To develop a set of open-source software tools for CT, proton, and He-3 pulmonary computational analysis. Development will include several basic tools

Atlas-based lung segmentation. Identification of anatomical structure in MRI is often a crucial preprocessing step for quantification of morphological features or functional information. Quantitative regional analysis often requires the identification of lung and lobar anatomy. Although much algorithmic research for lung segmentation has been reported in the CT literature [34], co-opting such technologies is complicated by MRI-specific issues such as RF coil inhomogeneity, presence and resolution of structural detail, and the absence of a physically-based intensity scaling.

We recently proposed a multi-atlas approach for automatically segmenting the left and right lungs in proton MRI [31]. Multi-atlas approaches to segmentation have proven highly successful in neuroimaging [Reference 11;Wang:2013aa] which translates readily to a pulmonary context. Many current strategies for lung image segmentation employ low-level processing techniques based on encodable heuristics. Consensus-based strategies, in contrast, optimize the prior knowledge applied to a specific segmentation problem.

B-spline-based Symmetric Normalization. A thorough comparative evaluation with the well-known ANTs SyN algorithm was performed with a B-spline variant. The evaluation utilized multiple publicly available, annotated brain data sets and demonstrated statistically significant improvement in label overlap measures [30]. We also used the EMPIRE10 challenge framework to provide an additional comparison in the context of pulmonary CT image registration [35]. Due to the performance of this new variant, it has become the preferred transformation model for small deformation image registration problems (e.g., lung and cardiac [36] applications).

Multi-feature CT and multi-modal MRI template generation. we generate subject-specific templates directly from the image data. Given the variability in lung shape across populations and the lack of publicly available lung atlases, generating population- or subject-specific templates enhances the accuracy of the longitudinal analysis described in this work. Applicable to pulmonary data is the template construction algorithm described in [7] which was applied to T1-weighted brain data. However, the simultaneous acquisition of the 3He and 1H images lends itself to multimodal processing [10] in which both modalities are used to simultaneously produce 3He and 1H templates. This process is represented in Fig. 1 for a single subject.

Atlas-based lobe estimation. For regional investigation of certain lung pathologies and conditions, it is often useful to quantify measurements of interest within more localized regions, such as the lobes. However, as mentioned previously, there is little (if any) usable information in proton MRI for image-based lobar segmentation which has led to alternative geometric subdivisions which are ad hoc, non-anatomical, and do not adequately address intra- and inter-subject correspondences. However, we can take advantage of inter-subject similarities in lobar geometry to provide a prior-based estimation of lobar divisions using a consensus labeling approach.

To generate the lobe segmentation in a target proton lung image, we first generate the binary lung mask for the proton lung image as described in the previous subsection. We then register the set of CT lung masks to the target binary lung mask using the same registration approach mentioned earlier [30]. Subsequently, we warp the set of CT lobe labels to the target image using the CT mask-to-proton mask transformation. Since we have no intensity information inside the target lung mask and CT atlas lung masks, we use a simply majority voting strategy to generate the optimal labeling for the target image. Following the majority voting, we remove any labelings outside the lung mask and assign any unlabeled voxels with the label closest in distance to that voxel.

Ventilation-based image segmentation. Developments in MRI research utilizing noble gases, such as 3He and 129Xe, have demonstrated the capability of visualizing alveolar and bronchial air spaces. Currently, hyperpolarized 3He MRI is a low-risk investigatory technique that provides high spatial and temporal resolution images of the air spaces of the lungs and has been used to investigate a variety of lung diseases. Automated or semiautomated approaches for classifying areas of varying degrees of ventilation are of potential benefit for facilitating such investigation.

Airway and vessel segmentation. We should propose to implement something here. We should look at the Slicer/VMTK airway segmentation model.

Specific Aim 2. To provide multiple sets of multi-modal annotated lung data (CT, proton, and He3) for public use.

Volumetric Tissue Indices

lung volume
lobar volume
surface area
surface area to volume ratio
total lung weight
tissue/airspace volumes of lung
inspiration vs. expiration*

Airway Indices

airway luminal diameter and area
airway wall thickness
percentage wall area
thickness to diameter ratio
airway branch angles
airway segment length
airway wall volumes (segmental and total)*
inspiration vs. expiration

Distribution of LAA Heterogeneity

10 partitions (std of $15^{th}\%$)
slopes of density mask curves
$\%$ size distribution of LAA areas
volumetric cluster analysis
inner core vs. outer rind
inspiration vs. expiration*

Cooccurrence Matrix Texture Indices

energy
inertia
contrast
entropy
correlation
inverse difference moment
cluster shade*
cluster prominence*
Haralick's correlation*

Run-length Matrix Texture Indices

short run emphasis
long run emphasis
grey level non-uniformity
run-length non-uniformity
run percentage
low grey level run emphasis*
high grey level run emphasis*
short run low grey level emphasis*
short high grey level run emphasis*
long run low grey level emphasis*
long high grey level run emphasis*
inspiration vs. expiration*

Attenuation Histogram Statistics

attenuation mean
attenuation variance
attenuation skewness
attenuation kurtosis
attenuation grey level entropy
regional variants
inspiration vs. expiration

Deformation Indices

Jacobian of lung displacement
lung deformation strain

Stochastic Fractal Image Statistics

mean
variance
skewness
kurtosis
grey level entropy
inspiration vs. expiration*

Attenuation Mask Indices

HU density mask
$\%$ HU density mask
inspiration vs. expiration*

Table 1: Quantitative CT indices proposed for inclusion in the lung image analysis pipeline. Whole lung, regional, and voxelwise measurements are included, as well as population-based comparisons and longitudinal analysis of all indices. Indices marked with a '*' denote novel measures which have not been previously utilized in chronic lung disease assessment but have shown classification capability in other application domains.

References

1. Fischl, B. “**FreeSurfer**” *Neuroimage* 62, no. 2 (2012): 774–81. doi:10.1016/j.neuroimage.2012.01.021

2. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. “**FSL**” *Neuroimage* 62, no. 2 (2012): 782–90. doi:10.1016/j.neuroimage.2011.09.015

3. Cox, R. W. “**AFNI: what a Long Strange Trip It’s Been**” *Neuroimage* 62, no. 2 (2012): 743–7. doi:10.1016/j.neuroimage.2011.08.

4. Ashburner, J. “**SPM: a History**” *Neuroimage* 62, no. 2 (2012): 791–800. doi:10.1016/j.neuroimage.2011.10.025

5. Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. “**A Reproducible Evaluation of ANTs Similarity Metric Performance in Brain Image Registration**” *Neuroimage* 54, no. 3 (2011): 2033–44. doi:10.1016/j.neuroimage.2010.09.025

6. Avants, B. B., Tustison, N. J., Stauffer, M., Song, G., Wu, B., and Gee, J. C. “**The Insight ToolKit Image Registration Framework**” *Front Neuroinform* 8, (2014): 44. doi:10.3389/fninf.2014.00044

7. Avants, B. B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., and Gee, J. C. “**The Optimal Template Effect in Hippocampus Studies of Diseased Populations**” *Neuroimage* 49, no. 3 (2010): 2457–66. doi:10.1016/j.neuroimage.2009.09.062

8. Avants, B. B., Klein, A., Tustison, N. J., Woo, J., and Gee, J. C. “**Evaluation of Open-Access, Automated Brain Extraction Methods on Multi-Site Multi-Disorder Data**” *16th annual meeting for the organization of human brain mapping* (2010):

9. Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., Kandel, B. M., Strien, N. van, Stone, J. R., Gee, J. C., and Avants, B. B. “**Large-Scale Evaluation of ANTs and FreeSurfer Cortical Thickness Measurements**” *Neuroimage* 99, (2014): 166–79. doi:10.1016/j.neuroimage.2014.05.044

10. Tustison, N. J., Shrinidhi, K. L., Wintermark, M., Durst, C. R., Kandel, B. M., Gee, J. C., Grossman, M. C., and Avants, B. B. “**Optimal Symmetric Multimodal Templates and Concatenated Random Forests for Supervised Brain Tumor Segmentation (Simplified) with** $ANTsR$” *Neuroinformatics* (2014): doi:10.1007/s12021-014-9245-2

11. Wang, H., Suh, J. W., Das, S. R., Pluta, J., Craige, C., and Yushkevich, P. A. “**Multi-Atlas Segmentation with Joint Label Fusion**” *IEEE Trans Pattern Anal Mach Intell* (2012): doi:10.1109/TPAMI.2012.143

12. Wang, H. and Yushkevich, P. A. “**Multi-Atlas Segmentation with Joint Label Fusion and Corrective Learning-an Open Source Implementation**” *Front Neuroinform* 7, (2013): 27. doi:10.3389/fninf.2013.00027

13. Menze, B., Reyes, M., and Van Leemput, K. “**The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)**” *IEEE Trans Med Imaging* (2014): doi:10.1109/TMI.2014.2377694

14. Sluimer, I., Schilham, A., Prokop, M., and Ginneken, B. van. “**Computer Analysis of Computed Tomography Scans of the Lung: a Survey**” *IEEE Trans Med Imaging* 25, no. 4 (2006): 385–405. doi:10.1109/TMI.2005.862753

15. De Nunzio, G., Tommasi, E., Agrusti, A., Cataldo, R., De Mitri, I., Favetta, M., Maglio, S., Massafra, A., Quarta, M., Torsello, M., Zecca, I., Bellotti, R., Tangaro, S., Calvini, P., Camarlinghi, N., Falaschi, F., Cerello, P., and Oliva, P. “**Automatic Lung Segmentation in CT Images with Accurate Handling of the Hilar Region**” *J Digit Imaging* 24, no. 1 (2011): 11–27. doi:10.1007/s10278-009-9229-1

16. Prasad, M. N., Brown, M. S., Ahmad, S., Abtin, F., Allen, J., Costa, I. da, Kim, H. J., McNitt-Gray, M. F., and Goldin, J. G. “**Automatic Segmentation of Lung Parenchyma in the Presence of Diseases Based on Curvature of Ribs**” *Acad Radiol* 15, no. 9 (2008): 1173–80. doi:10.1016/j.acra.2008.02.004

17. Wang, J., Li, F., and Li, Q. “**Automated Segmentation of Lungs with Severe Interstitial Lung Disease in CT**” *Med Phys* 36, no. 10 (2009): 4592–9.

18. Rikxoort, E. M. van, Hoop, B. de, Viergever, M. A., Prokop, M., and Ginneken, B. van. “**Automatic Lung Segmentation from Thoracic Computed Tomography Scans Using a Hybrid Approach with Error Detection**” *Med Phys* 36, no. 7 (2009): 2934–47.

19. Zheng, B., Leader, J. K., McMurray, J. M., Park, S. C., Fuhrman, C. R., Gur, D., and Sciurba, F. C. "**Automated Detection and Quantitative Assessment of Pulmonary Airways Depicted on CT Images**" *Med Phys* 34, no. 7 (2007): 2844–52.

20. Nakamura, M., Wada, S., Miki, T., Shimada, Y., Suda, Y., and Tamura, G. "**Automated Segmentation and Morphometric Analysis of the Human Airway Tree from Multidetector CT Images**" *J Physiol Sci* 58, no. 7 (2008): 493–8. doi:10.2170/physiolsci.RP007408

21. Lo, P., Ginneken, B. van, Reinhardt, J. M., and Bruijne, M. de. "**Extraction of Airways from CT (EXACT '09)**" *The second international workshop on pulmonary image analysis* (2009):

22. Agam, G., Armato, S. G., 3rd, and Wu, C. "**Vessel Tree Reconstruction in Thoracic CT Scans with Application to Nodule Detection**" *IEEE Trans Med Imaging* 24, no. 4 (2005): 486–99.

23. Korfiatis, P. D., Kalogeropoulou, C., Karahaliou, A. N., Kazantzi, A. D., and Costaridou, L. I. "**Vessel Tree Segmentation in Presence of Interstitial Lung Disease in MDCT**" *IEEE Trans Inf Technol Biomed* 15, no. 2 (2011): 214–20. doi:10.1109/TITB.2011.2112668

24. Qi, S., Triest, H. J. W. van, Yue, Y., Xu, M., and Kang, Y. "**Automatic Pulmonary Fissure Detection and Lobe Segmentation in CT Chest Images**" *Biomed Eng Online* 13, (2014): 59. doi:10.1186/1475-925X-13-59

25. Doel, T., Gavaghan, D. J., and Grau, V. "**Review of Automatic Pulmonary Lobe Segmentation Methods from CT**" *Comput Med Imaging Graph* 40, (2015): 13–29. doi:10.1016/j.compmedimag.2014.10.008

26. Uppaluri, R., Hoffman, E. A., Sonka, M., Hartley, P. G., Hunninghake, G. W., and McLennan, G. "**Computer Recognition of Regional Lung Disease Patterns**" *Am J Respir Crit Care Med* 160, no. 2 (1999): 648–54. doi:10.1164/ajrccm.160.2.9804094

27. Rosas, I. O., Yao, J., Avila, N. A., Chow, C. K., Gahl, W. A., and Gochuico, B. R. "**Automated Quantification of High-Resolution CT Scan Findings in Individuals at Risk for Pulmonary Fibrosis**" *Chest* 140, no. 6 (2011): 1590–7. doi:10.1378/chest.10-2545

28. DeBoer, E. M., Swiercz, W., Heltshe, S. L., Anthony, M. M., Szefler, P., Klein, R., Strain, J., Brody, A. S., and Sagel, S. D. "**Automated CT Scan Scores of Bronchiectasis and Air Trapping in Cystic Fibrosis**" *Chest* 145, no. 3 (2014): 593–603. doi:10.1378/chest.13-0588

29. Tustison, N. J., Avants, B. B., Flors, L., Altes, T. A., Lange, E. E. de, Mugler, J. P., 3rd, and Gee, J. C. "**Ventilation-Based Segmentation of the Lungs Using Hyperpolarized (3)He MRI**" *J Magn Reson Imaging* 34, no. 4 (2011): 831–41. doi:10.1002/jmri.22738

30. Tustison, N. J. and Avants, B. B. "**Explicit B-Spline Regularization in Diffeomorphic Image Registration**" *Front Neuroinform* 7, (2013): 39. doi:10.3389/fninf.2013.00039

31. Tustison, N. J., Qing, K., Wang, C., Altes, T. A., and Mugler, J. P., 3rd. "**Atlas-Based Estimation of Lung and Lobar Anatomy in Proton MRI**" *Magn Reson Med* (Accepted):

32. Yunwen, Y. and Kishida, K. "**Toward an Understanding of the Motivation of Open Source Software Developers**" *Software engineering, 2003. proceedings. 25th international conference on* (2003): 419–429. doi:10.1109/ICSE.2003.1201220

33. (2008): Available at http://fsmsh.com/2845

34. Rikxoort, E. M. van and Ginneken, B. van. "**Automated Segmentation of Pulmonary Structures in Thoracic Computed Tomography Scans: a Review**" *Phys Med Biol* 58, no. 17 (2013): R187–220. doi:10.1088/0031-9155/58/17/R187

35. Tustison, N. J., Song, G., Gee, James C, and Avants, B. B. "**Two Greedy SyN Variants for Pulmonary Image Registration**" *Evaluation of methods for pulmonary image registration (EMPIRE10)* (2012):

36. Tustison, N. J., Yang, Y., and Salerno, M. "**Advanced Normalization Tools for Cardiac Motion Correction**" *Statistical atlases and computational models of the heart - imaging and modelling challenges* 8896, (2015): 3–12. doi:10.1007/978-3-319-14678-2_1, Available at http://dx.doi.org/10.1007/978-3-319-14678-2_1