

PDF Annotation and Data Harmonization (PANDHA)

September 2021
Version 1.2

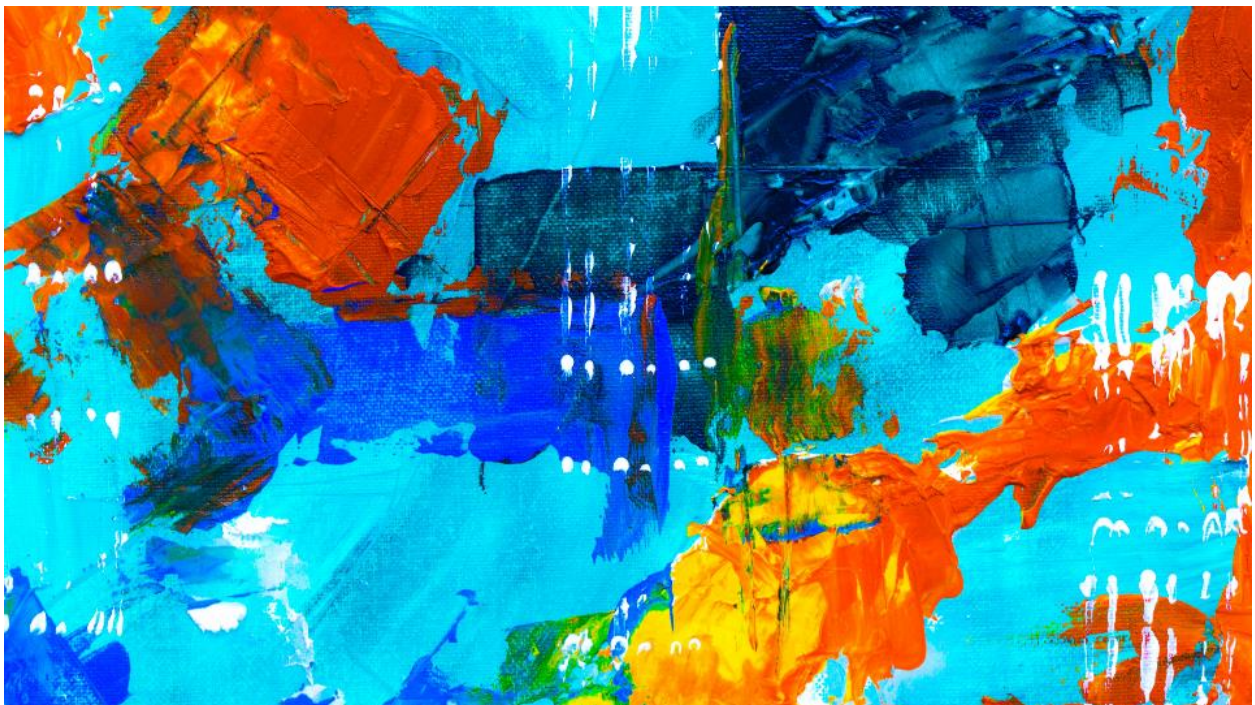


Table of Contents

| | |
|--|----|
| Table of Contents | 1 |
| Document Revision | 2 |
| 1 Installation | 3 |
| 1.1 Scope and Purpose | 3 |
| 1.2 Source Code Access | 3 |
| 1.2 Installation MacOSX * need to work out zip issue | 3 |
| 1.3 Installation Windows | 4 |
| 2 Application | 8 |
| 2.0 Load PDF | 8 |
| 2.1 Annotate PDF | 10 |
| 2.2 Save Annotations | 12 |
| 2.3 Load Annotations | 13 |

Document Revision

| Date | Version Number | Document Changes |
|------------|----------------|------------------------------|
| 01/05/2021 | 1.0 | Initial draft |
| 07/20/2021 | 1.1 | Updated to beta PANDHA V1 |
| 09/29/2021 | 1.2 | Update to beta PANDHA V1.0.2 |

| | | |
|--|--|--|
| | | |
|--|--|--|

1 Installation

1.1 Scope and Purpose

The purpose of this project is to further the research and development of tools that NCEA can use in their creation of machine-readable datasets and machine learning research. This effort consists of the following objectives:

- Research and develop software for NCEA that provides the ability to annotate scientific publications for use in machine learning algorithms. This software should be able to accept a list of tags provided by NCEA, allow the user to apply these tags to PDF documents in a web interface, and then extract out the information needed in machine-readable formats that can be used for machine learning.

1.2 Source Code Access

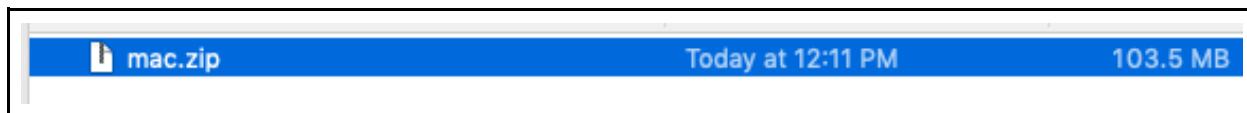
This describes the process of accessing the source code.

1. <https://github.com/USEPA/pdf-data-extraction>

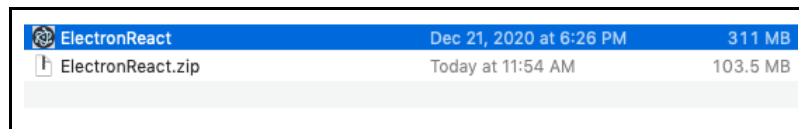
1.2 Installation MacOSX * need to work out zip issue

1. Download latest version from

2. Double click 'mac.zip' to unzip the file.



3. Double click 'ElectronReact' to launch the application.

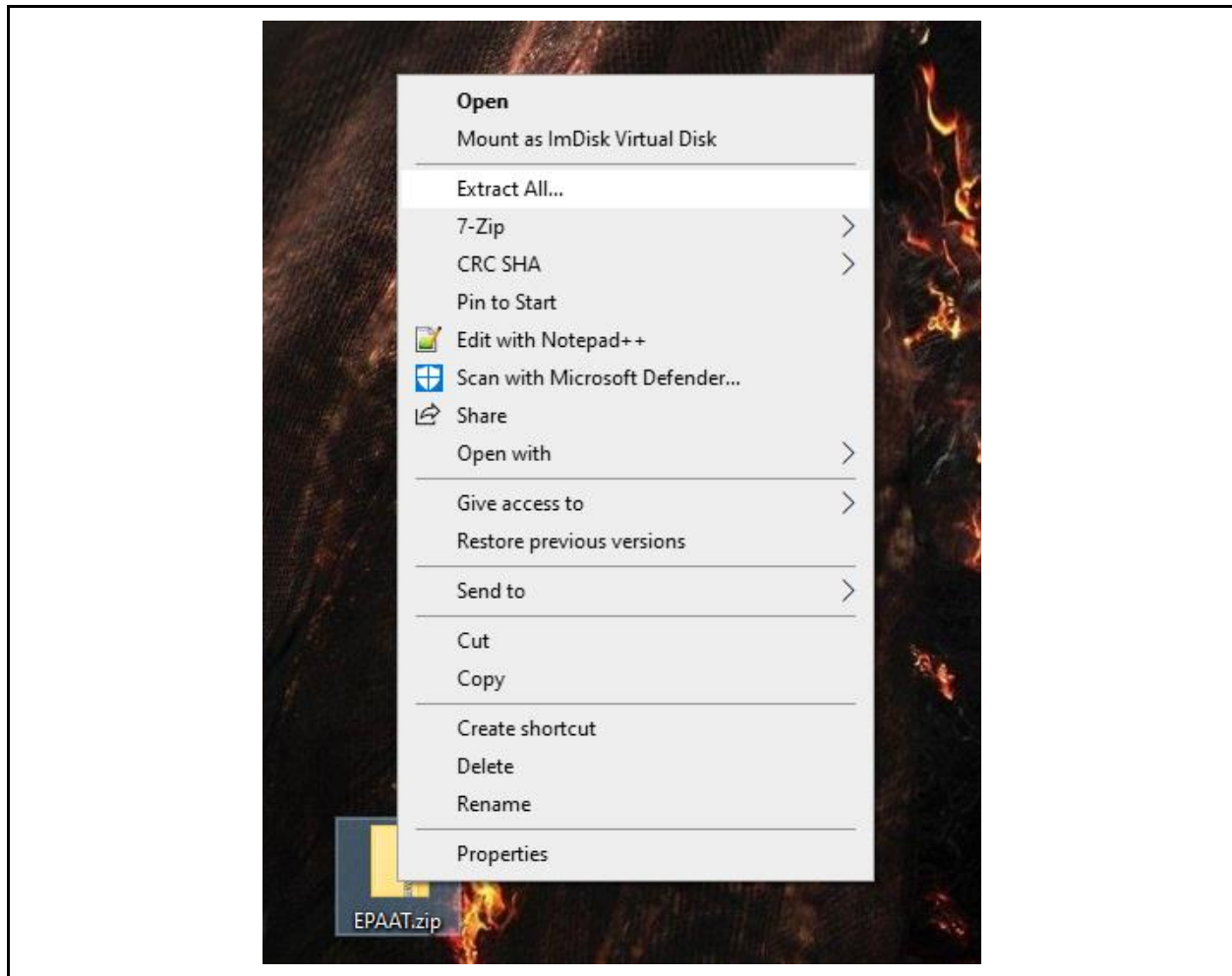


1.3 Installation Windows

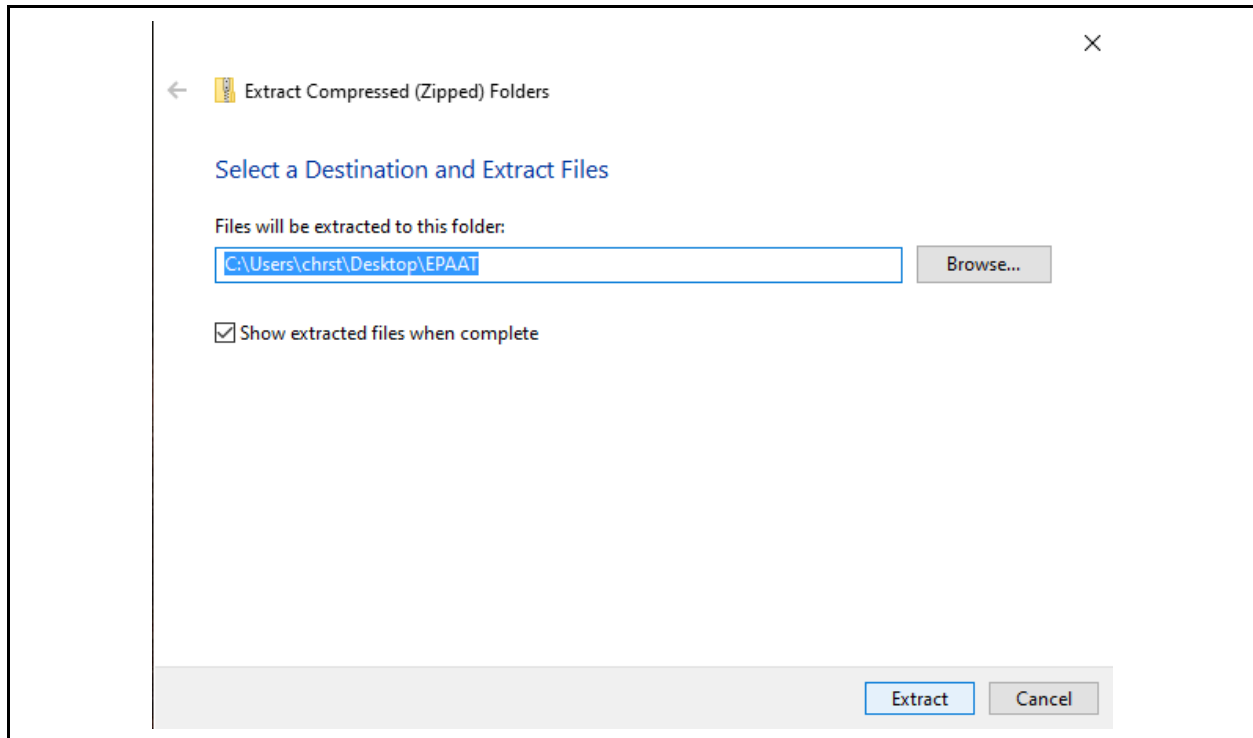
1. Download latest version from (PANDHA_1.0.2)

<https://usepa.sharepoint.com/sites/DOEORNLCPAD/Shared%20Documents/General/Annotation%20Tool%20Beta/Windows>

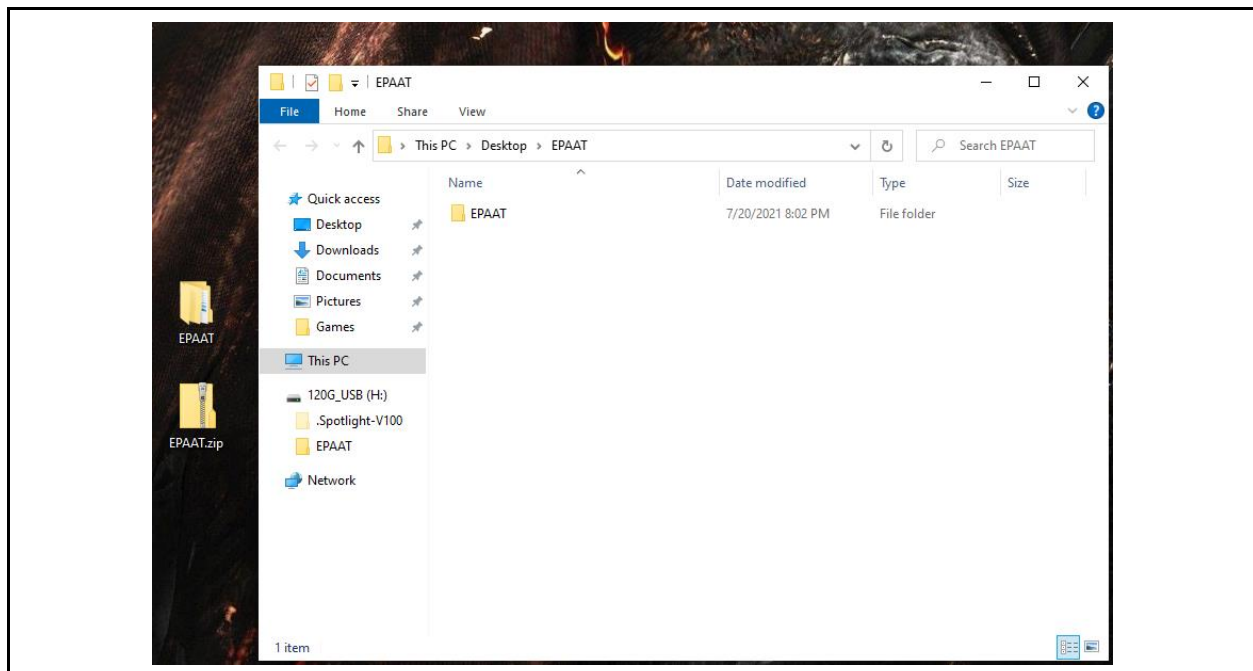
2. Right click 'Windows.zip' and select 'Extract All'



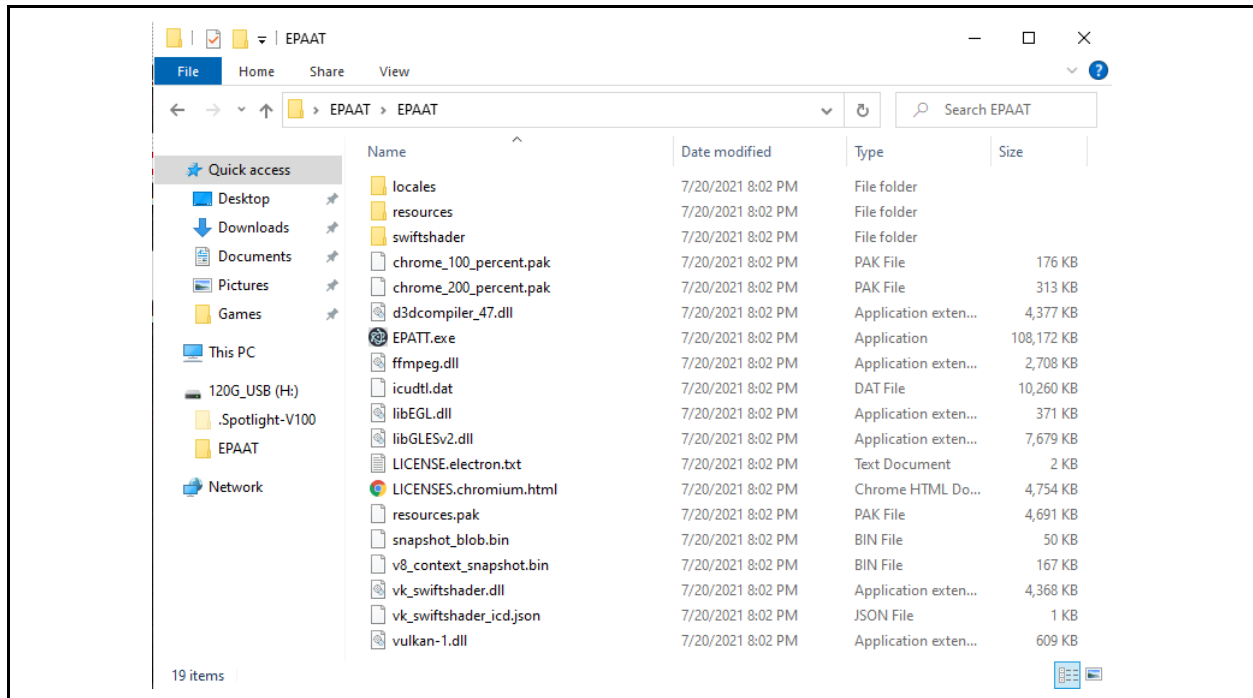
3. Select location and hit *Extract*



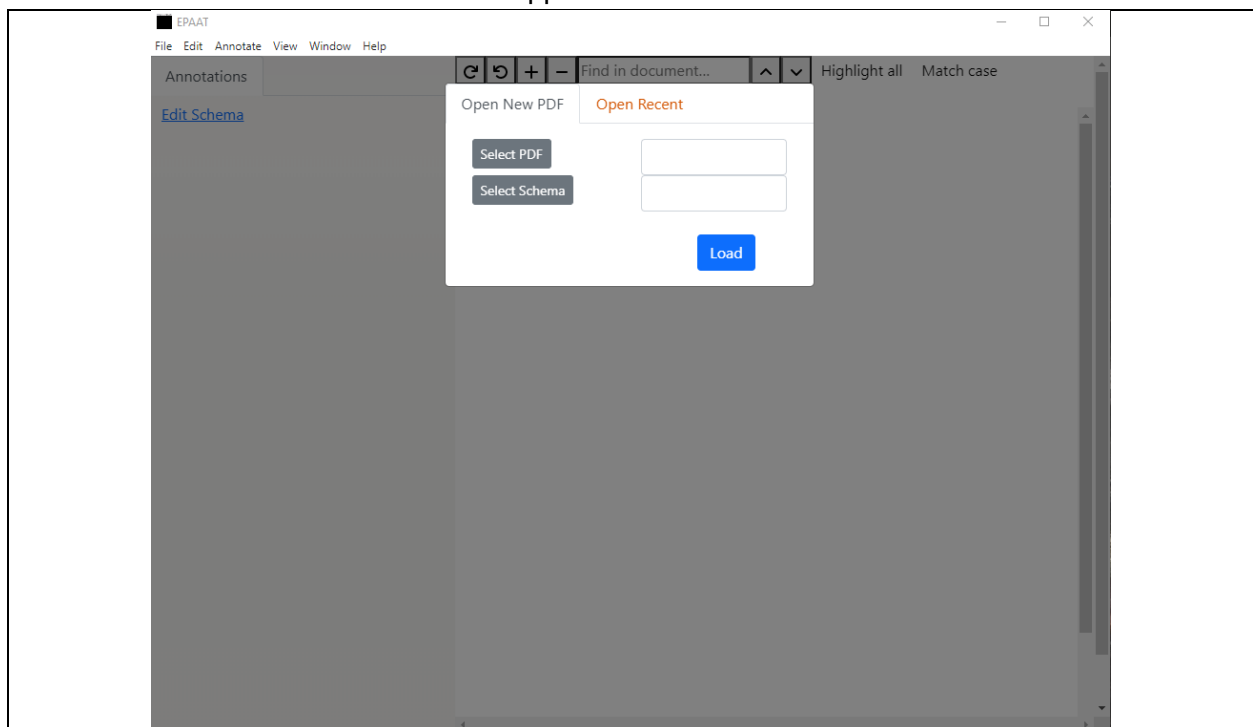
4. From the extracted location double click the *EPAAT* Folder



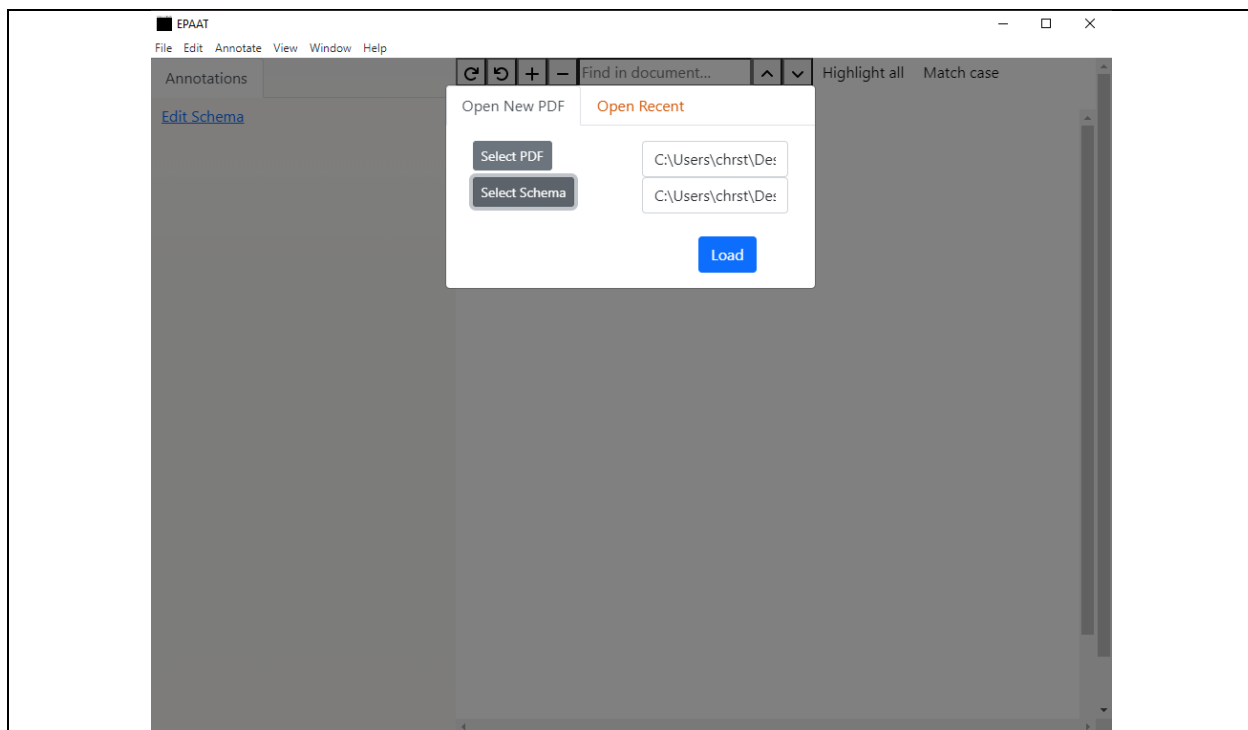
5. Double click *EPAAT*



6. Double click *EPAAT.exe* to start the application



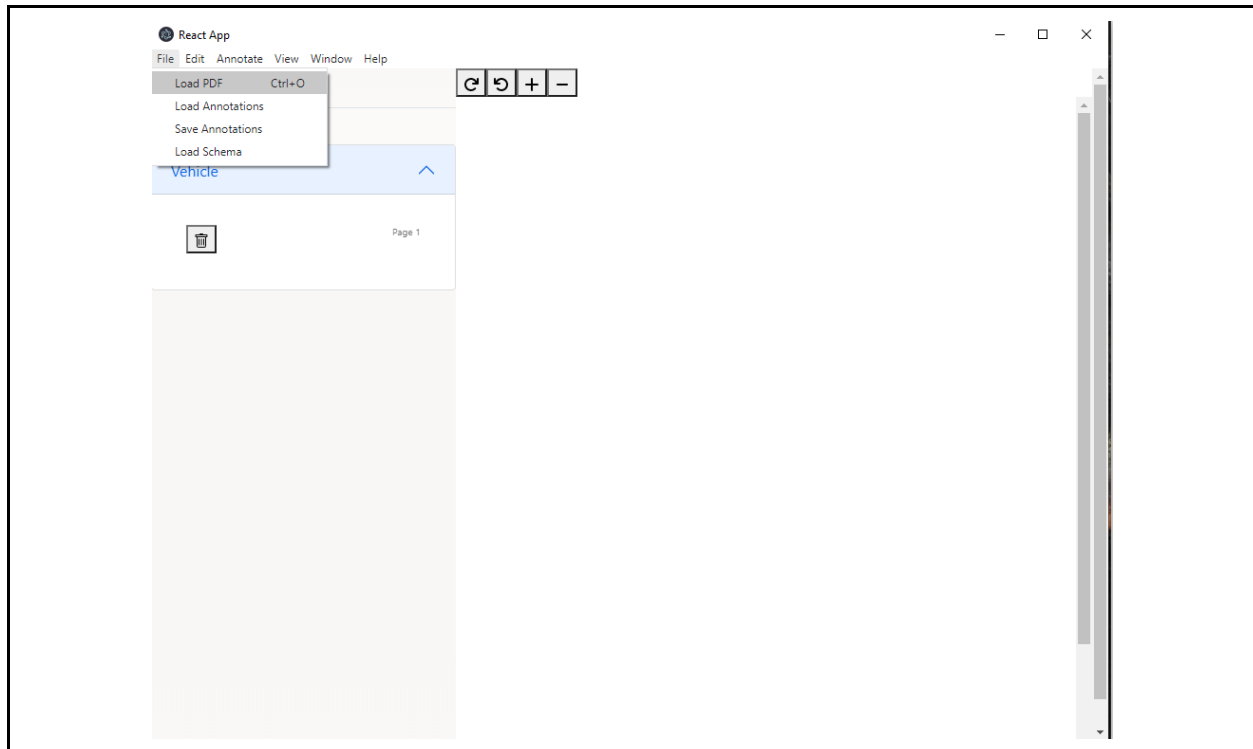
7. Select the PDF and Schema (schema.json is including in the installation folder) and click *Load*



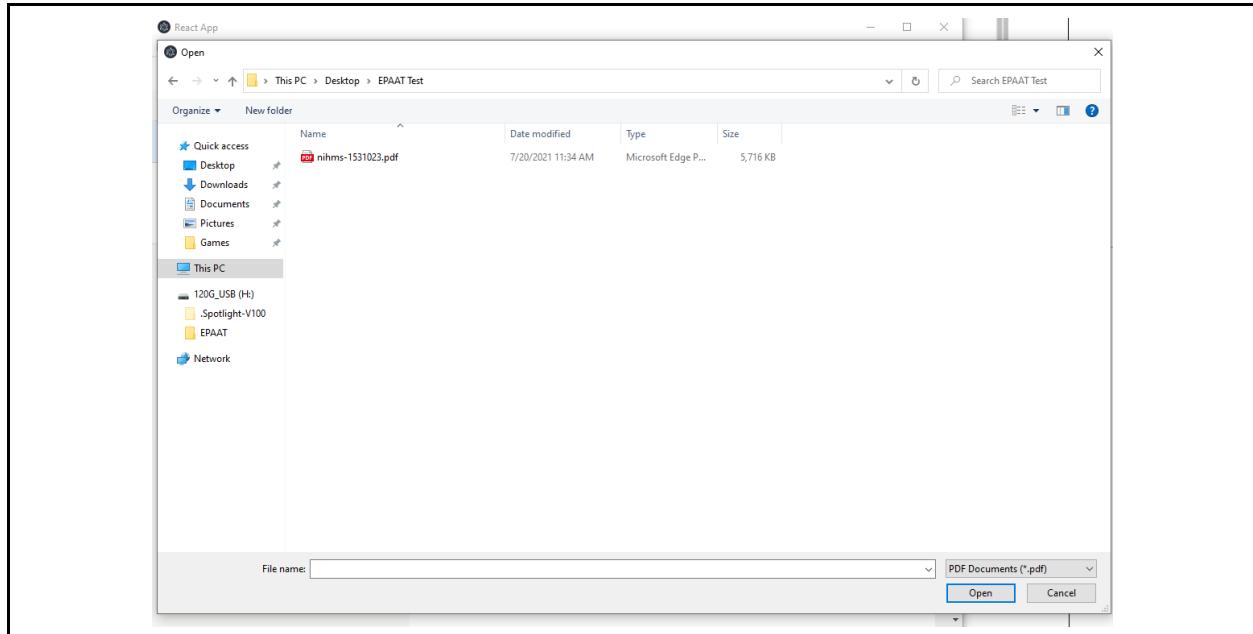
2 Application

2.0 Load PDF

1. Click *File* in the menu bar and select *Load PDF*.



2. Select the PDF file from your computer.



React App

File Edit Annotate View Window Help

Annotations

Edit Schema

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript
Nature. Author manuscript; available in PMC 2020 November 28.

Published in final edited form as:
Nature. 2019 July ; 571(7764): 270–274. doi:10.1038/s41586-019-1324-y.

TOX is a critical regulator of tumour-specific T cell differentiation

Andrew C. Scott^{1,2}, Friederike Dündar^{3,4}, Paul Zumbo^{3,4}, Smita S. Chandran^{5,6}, Christopher A. Klebanoff^{6,7,8}, Mojdeh Shakiba^{1,3}, Prerak Trivedi¹, Laura Menocal^{1,2}, Heather Appleby¹, Steven Camara¹, Dmitriy Zamarin^{6,7}, Tyler Walther⁷, Alexandra Snyder⁷, Matthew R. Femia^{6,9}, Elizabeth A. Comen^{7,8}, Hannah Y. Wen⁹, Matthew D. Hellmann^{5,7,8}, Niroshana Anandasabapathy^{2,10}, Yong Liu¹⁰, Nasser K. Altorki¹¹, Peter Lauer¹², Olivier Levy¹, Michael S. Glickman^{1,2}, Jonathan Kaye¹³, Doron Betel^{4,14,15}, Mary Philip^{1,16,*}, Andrea Schietinger^{1,2,5,*}

¹Immunology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
²Immunology and Microbial Pathogenesis Program, Weill Cornell Graduate School of Medical Sciences, New York, NY, USA.
³Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA.
⁴Applied Bioinformatics Core, Weill Cornell Medicine, New York, NY, USA.
⁵Parker Institute for Cancer Immunotherapy, New York, NY, USA.
⁶Center for Cell Engineering, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
⁷Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
⁸Weill Cornell Medical College, New York, NY, USA.
⁹Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
¹⁰Department of Dermatology, Weill Cornell Medical College, New York, NY, USA.

*Correspondence and requests for materials should be addressed to M.P. and A. Schietinger. mary.philip@mskcc.org; aschietinger@mskcc.org

Author contribution: A.C.S., M.P. and A. Schietinger conceived and designed the study. A.C.S., M.P., D.B., F.D., P.Z. and A. Schietinger conceived the computational analyses. D.B., F.D. and P.Z. performed all of the computational analyses. A.C.S., M.P., P.T., L.M., H.Y., H.A. and S.S.C. carried out experiments. A.C.S., M.P., F.D., P.Z., D.B., S.S.C., C.A.K. and A. Schietinger interpreted data. S.C. and H.A. assisted with mouse breeding. T.W., A. Snyder, D.Z., M.D.H., M.R.F., E.A.C., H.Y.W. and C.A.K. provided human samples. S.A., Y.L. and N.K.A. contributed to the analysis of human samples. O.L. and M.S.G. provided help in establishing the knockout model. O.L., M.S.G. and J.K. provided mice. P.L. provided *LoxP* strains. A.C.S., M.P., F.D., P.Z., D.B. and A. Schietinger wrote the manuscript, with all authors contributing to writing and providing feedback.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information, details of author contributions and competing interests, and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1324-y>.

Data availability
All data generated and supporting the findings of this study are available within the paper. The RNA-seq and ATAC-seq data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE126974. Source Data are provided with the online version of the paper. Additional information and materials will be made available upon request.

Additional information
Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1324-y>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

2.1 Annotate PDF

1. Highlight text you wish to annotate and select *Add Annotation*.

React App

File Edit Annotate View Window Help

Annotations

Edit Schema

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript
Nature. Author manuscript; available in PMC 2020 November 28.

Published in final edited form as:
Nature. 2019 July ; 571(7764): 270–274. doi:10.1038/s41586-019-1324-y.

TOX is a critical regulator of tumour-specific T cell differentiation

Andrew C. Scott^{1,2}, Friederike Dündar^{3,4}, Paul Zumbo^{3,4}, Smita S. Chandran^{5,6}, Christopher A. Klebanoff^{6,7,8}, Mojdeh Shakiba^{1,3}, Prerak Trivedi¹, Laura Menocal^{1,2}, Heather Appleby¹, Steven Camara¹, Dmitriy Zamarin^{6,7}, Tyler Walther⁷, Alexandra Snyder⁷, Matthew R. Femia^{6,9}, Elizabeth A. Comen^{7,8}, Hannah Y. Wen⁹, Matthew D. Hellmann^{5,7,8}, Niroshana Anandasabapathy^{2,10}, Yong Liu¹⁰, Nasser K. Altorki¹¹, Peter Lauer¹², Olivier Levy¹, Michael S. Glickman^{1,2}, Jonathan Kaye¹³, Doron Betel^{4,14,15}, Mary Philip^{1,16,*}, Andrea Schietinger^{1,2,5,*}

¹Immunology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
²Immunology and Microbial Pathogenesis Program, Weill Cornell Graduate School of Medical Sciences, New York, NY, USA.
³Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA.
⁴Applied Bioinformatics Core, Weill Cornell Medicine, New York, NY, USA.
⁵Parker Institute for Cancer Immunotherapy, New York, NY, USA.
⁶Center for Cell Engineering, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
⁷Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
⁸Weill Cornell Medical College, New York, NY, USA.
⁹Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
¹⁰Department of Dermatology, Weill Cornell Medical College, New York, NY, USA.

*Correspondence and requests for materials should be addressed to M.P. and A. Schietinger. mary.philip@mskcc.org; aschietinger@mskcc.org

Author contribution: A.C.S., M.P. and A. Schietinger conceived and designed the study. A.C.S., M.P., D.B., F.D., P.Z. and A. Schietinger conceived the computational analyses. D.B., F.D. and P.Z. performed all of the computational analyses. A.C.S., M.P., P.T., L.M., H.Y., H.A. and S.S.C. carried out experiments. A.C.S., M.P., F.D., P.Z., D.B., S.S.C., C.A.K. and A. Schietinger interpreted data. S.C. and H.A. assisted with mouse breeding. T.W., A. Snyder, D.Z., M.D.H., M.R.F., E.A.C., H.Y.W. and C.A.K. provided human samples. S.A., Y.L. and N.K.A. contributed to the analysis of human samples. O.L. and M.S.G. provided help in establishing the knockout model. O.L., M.S.G. and J.K. provided mice. P.L. provided *LoxP* strains. A.C.S., M.P., F.D., P.Z., D.B. and A. Schietinger wrote the manuscript, with all authors contributing to writing and providing feedback.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information, details of author contributions and competing interests, and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1324-y>.

Data availability
All data generated and supporting the findings of this study are available within the paper. The RNA-seq and ATAC-seq data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE126974. Source Data are provided with the online version of the paper. Additional information and materials will be made available upon request.

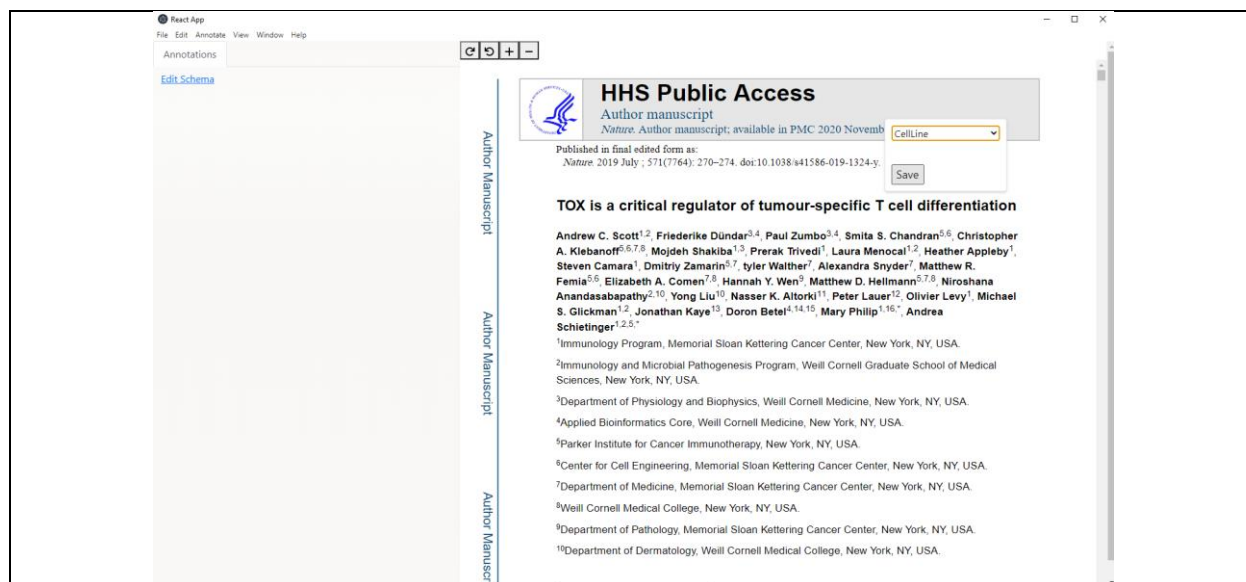
Additional information
Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1324-y>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

2. Select the annotation type.



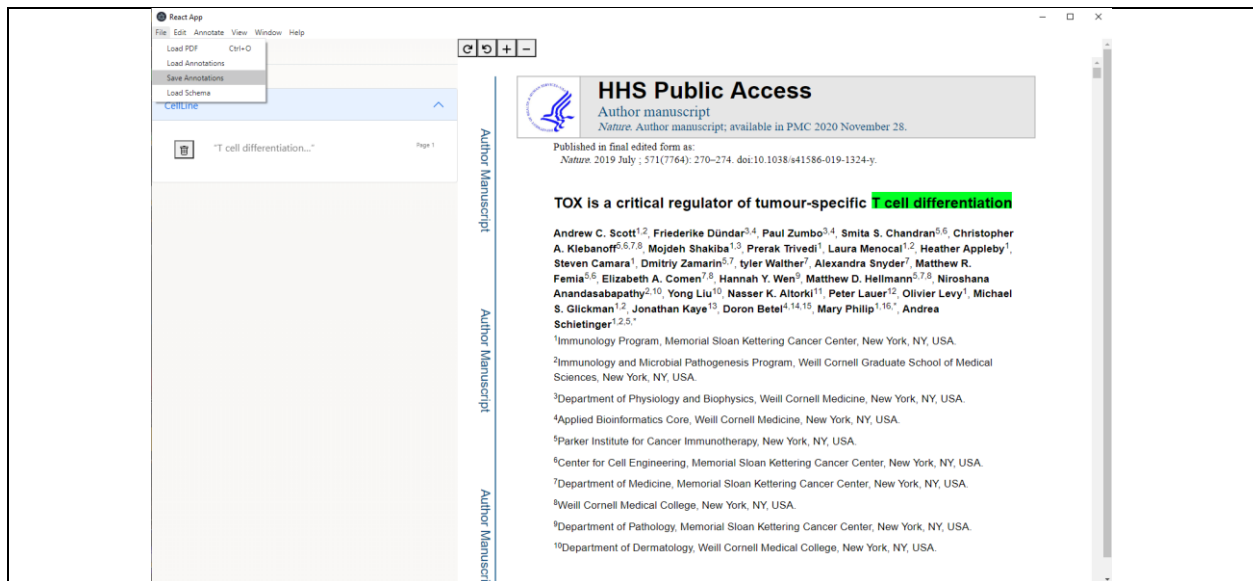
3. Hit save





2.2 Save Annotations

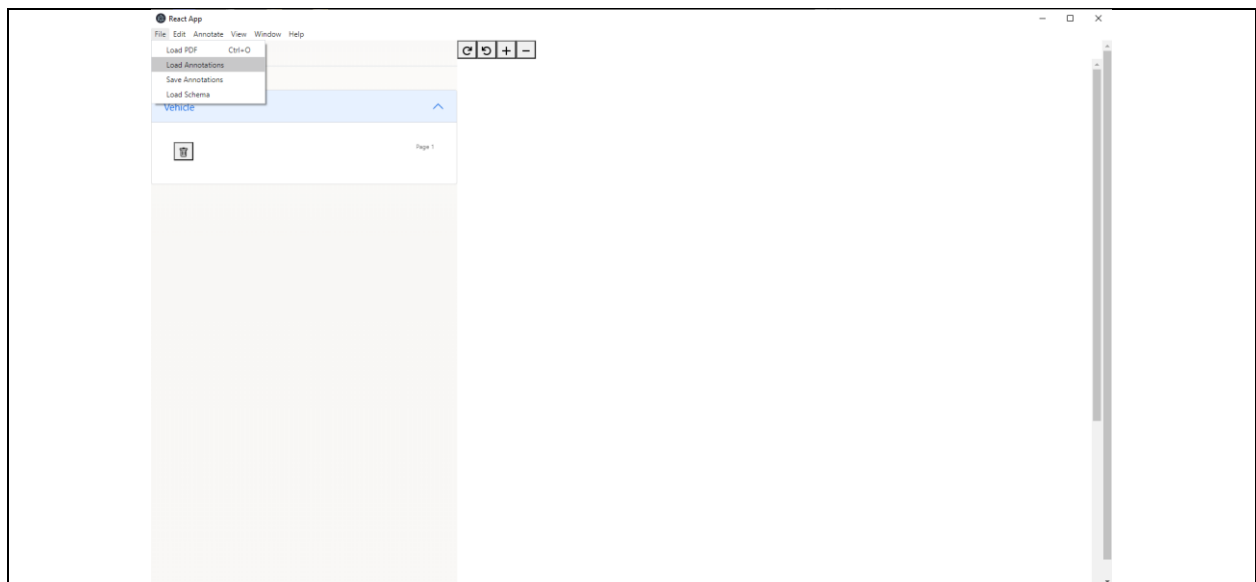
1. Click *File* in the menu bar and select *Save Annotations*.



2. Select a save location on your computer and click *Save Annot File*.

1. Click *File* in the menu bar and select *Load Annotations*.

1. Click *File* in the menu bar and select *Load Annotations*.



- ## 2. Select an annotation file

