UC Riverside
UC Riverside Electronic Theses and Dissertations

Title
Examining NeuroHIV-Induced Immune Responses Through Gene Expression and a Newly Designed Machine Learning-Based Tool for Automatic Image Analysis

Permalink
https://escholarship.org/uc/item/9gb0p3gh

Author
Kataras, Theodore John

Publication Date
2022

Copyright Information
This work is made available under the terms of a Creative Commons Attribution-ShareAlike License, availalbe at https://creativecommons.org/licenses/by-sa/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Examining NeuroHIV-Induced Immune Responses Through Gene Expression and a Newly Designed Machine Learning-Based Tool for Automatic Image Analysis

A Dissertation Submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Genetics, Genomics and Bioinformatics

by

Theodore John Kataras

December 2022

Dissertation Committee:
Dr. Marcus Kaul, Chairperson
Dr. Wenxiu Ma
Dr. Adam Godzik

Copyright by
Theodore John Kataras
2022

The Dissertation of Theodore John Kataras is approved:

Committee Chairperson

University of California, Riverside

Acknowledgements

All work was performed under the guidance of Dr. Marcus Kaul at the University of California Riverside, Riverside, CA (UCR). The gene expression microarray data analysis that formed the basis for the identification of guanine-binding proteins as factors of interest was performed by Dr. Roy Williams (Sanford Burnham Medical Research Institute, La Jolla, CA) in collaboration with Dr. Kaul. Monocyte derived macrophage culture and HIV infection performed by Dr. Ana Sanchez (Sanford Burnham Medical Research Institute, La Jolla, CA) and the RNA Seq data analysis was performed by Cory White (University of California San Diego, La Jolla, CA). Population Specific Enrichment Analysis strategy introduced by Dr. Lukasz Jaroszewski. Automatic cell counting work and code co-authored with Tyler Jang. Dr. Nina Yuan, Dr. Hina Singh and Dr. Deepika Bhullar for images of mouse brain tissues .Hand-counting of image data sets was assisted by Dr. Hina Singh and Dominic Fok. Guidance on experiments and data analysis was provided by Dr. Hina Singh and Jeffrey Koury.

ABSTRACT OF THE DISSERTATION

Examining NeuroHIV-Induced Immune Responses Through Gene Expression and a Newly Designed Machine Learning-Based Tool for Automatic Image Analysis

by

Theodore John Kataras

Doctor of Philosophy, Graduate Program in Genetics
Genomics and Bioinformatics.
University of California, Riverside, December 2022
Dr. Marcus Kaul, Chairperson

Human Immunodeficiency Virus (HIV) remains a global health concern even as antiretroviral therapies (ART) improve livelihood outcomes for infected individuals. Viral infection in the brain presents a reservoir removed from the effects of these treatments. As the population of HIV infected individuals ages with greater longevity, neurocognitive decline has remained prevalent despite ART. Addressing this neurocognitive decline requires altering the chronic immuno-activation that appears to drive the dysfunction and developing interventions within a complete understanding of immune pathways involved.

This dissertation seeks to provide novel insight into genes involved in HIV induced neurodegeneration as well as introducing a program designed to enhance the throughput of this and similar studies on dynamic cellular systems. ACCT: Automatic Cell Counting with Trainable Weka Segmentation was developed to be accessible in both computational and computer science

expertise requirements and we have compared its performance to similar programs on neurological datasets. While we developed this program using data from neuroscience, it stands as a broadly accessible and useful tool across any imaging data with sufficiently defined objects to be quantified.

The primary resident immune cells of the brain are the microglia. These cells champion the innate immune response to viruses via the type 1 and 2 interferon response. Interferon response genes (IRG), however, also have the capability of feeding back on the pathway itself to modify the throughput of the immune response. Two genes which have regulatory effects on type 1 interferon response are Interferon Regulatory Factor 7 (IRF7), a promoter of induction and enhancement of type 1 interferon activity and Guanylate Binding Protein 4 (GBP4), an inhibitor of IRF7 function.

I have associated the ortholog to this GBP gene with immune response to HIV in an HIV infection model using gene expression data with population specific enrichment analysis (PSEA). I have also examined this protein's localization in mouse cortex, and observed distinct staining populations of immune cells. This dissertation seeks to advance understanding through increased understanding of HIV-induced neurodegeneration. The open-access design of ACCT is intended to increase researchers' ability to perform similar analyses using this computational tool that we have designed and contextualized within and against similar automatic image quantification methods.

Table of Contents:

Abstract *iv*

Introduction 1

Examining GBP involvement in HIV induced neurodegeneration 10

ACCT: a Fast and Accessible Automatic Cell Counting Tool Using Machine Learning for 2D Image Segmentation 37

Conclusion 67

List of Figures:

Introduction:

Fig. 1 IRF7 and GBP4 in Type I Interferon Response to Virus 3

Examining GBP involvement in HIV induced neurodegeneration:

Fig. 1 Correlation coefficient plot 16

Fig. 2 Explanation of PSEA process on mouse brain microarray data 17

Fig. 3 Guanylate Binding Proteins respond to HIV in immune cells inside and outside the brain 22

Fig. 4 GBP4 and IRF7 in human macrophage culture 23

Fig. 5 Gbp3 antibody in Activated mouse cortex at 10x magnification 24

Fig. 6 Gbp3 antibody titration in mouse cortex. 25

Fig. 7 Iba-1 and P2ry12 co-staining in mouse cortex 26

Fig. 8 Tomato lectin with Iba-1 or Gbp3 staining in mouse cortex 27

ACCT: a Fast and Accessible Automatic Cell Counting Tool Using Machine Learning for 2D Image Segmentation:

Fig. 1 A visual overview of ACCT components and process 43

Fig. 2 Images of immunofluorescence-labeled microglia with segmentation 45

Fig. 3 Image of fluorescence-labeled neurons with segmentation 47

Fig. 4 Summary of individual classifier performance on the Iba-1 microglia dataset during the validation stage 52

Fig. 5 Mean microglia density by experimental genotype in manual and automated counts 56

Fig. 6 ACCT vs Ilastik vs CellProfiler vs Manual Fiji on images of Iba-1 positive microglia 58

Fig. 7 ACCT vs c-ResUnet vs Ilastik on the Fluocell dataset 59

Fig. 8 An ROC curve generated by ACCT following the validation stage on Iba-1 microglia images 60

Figure S1. Map of the files and folders included in the ACCT pipeline 68

Figure S2. A comparison of the differences in images contained in the Fluocell dataset 69

Figure S3. Scatter plots with repression lines for all correlative comparisons between observer and automatics counts 70

Figure S4. ACCT applied on image of aerial photography 71

List of Tables:

ACCT: a Fast and Accessible Automatic Cell Counting Tool Using Machine Learning for 2D Image Segmentation:

Table 1. Correlation analysis of microglia density 55

Supplementary Table T1. Top 10 classifiers on Iba-1 validation data 72

Supplementary table T2. This is an example of the per cell morphological measurements recorded by ACCT on a single image in the Fluocell dataset 73

List of Abbreviations:

University of California, Riverside (UCR); Human Immunodeficiency Virus (HIV); antiretroviral therapies (ART); Interferon response genes (IRG); Interferon Regulatory Factor 7 (IRF7); Guanylate Binding Protein 4 (GBP4); Trainable Weka Segmentation (TWS); phosphate buffered saline (PBS); guanylate binding protein (GBP); Interferon (IFN); HIV brain infection (NeuroHIV); Acquired Immunodeficiency Syndrome (AIDS); central nervous system (CNS); reactive oxygen species (ROS); transgenic (tg); glial fibrillary acidic protein (GFAP); CCR5 knockout mice (CCR5KO); Immunoglobulin G (IgG); Population Specific Enrichment Analysis (PSEA); monocyte-derived macrophage (MDM); white blood cell concentrates (WBC); lipopolysaccharides (LPS); Automatic Cell Counting with Trainable Weka Segmentation (ACCT); graphical user interface (GUI); human immunodeficiency virus-1 (HIV-1); ionized calcium-binding adaptor protein-1 (Iba-1); convolutional neural net approach (CNN); mouse model for HIV-induced brain injury (HIVgp120tg); Institutional Animal Care and Use Committees (IACUC); Sanford Burnham Prebys Medical Discovery Institute (SBP); The Scripps Research Institute (TSRI); Fluorescein isothiocyanate (FITC); Receiver Operator Characteristic (ROC); mean absolute error (MAE); mean paired error (MPE);

Introduction

HIV causes mortality through a systemic drop in broad populations of immune cells due to chronic activation of activated CD4+ lymphocytes and macrophages. HIV additionally leads to a cascading immune response where the primary driver of dysfunction is not the infection itself, but the resulting chronic immune activation (Kaul et al., 2001; Laurence, 1993).

Advances in antiviral therapies have been successful in controlling the systemic HIV infection and massively improving the livelihood of individuals living with the virus. While these individuals live longer, it has become clear that HIV makes its way early during infection into the brain and presents different challenges to limit infection in the body at large (Kanmogne et al., 2020; Kaul et al., 2001).

The developed antiviral treatments for HIV have massively limited penetration beyond the blood-brain barrier. Because of this separation, it may be more effective to control HIV in the brain by altering the relationship between the existing immune mechanisms in the brain which drive dysfunction in response to HIV (Grant et al., 1999; Saloner & Cysique, 2017).

The immune landscape of the brain is unique behind the blood-brain barrier. The primary form of immune response in the brain relies on ubiquitous gene expression channels present in all cells, as well as a unique immune cell population. These unique immune cells are called microglia, and they are drivers of the innate immune response which targets broad types of

biological threats via an assortment of receptors that recognize elements common to viruses, bacteria, and other pathogens (D'Agostino et al., 2012; Gomez-Nicola & Perry, 2015).

The innate immune response to HIV and other immune threats is primarily comprised of ISGs (interferon stimulated genes). There are two type I interferon genes that transduce varied signaling and cytokine pathways in different manners. In the context of HIV, interferon alpha of type I interferon response has been found to be primarily responsible for the chronic immune activation that brings neurodegeneration, while interferon beta has been indicated to have a role in controlling the innate immune response to HIV (Savan, 2014; Singh et al., 2020; Thaney et al., 2017; Thaney & Kaul, 2019).

The interferon response is controlled through a network of transducers and repressors. IRF7 is an ISG which controls induction and continued transduction of the type I IFN pathway. This gene promotes initial activation, but also increases the intracellular availability of signaling molecules leading to further increased ISG activation (Honda et al., 2005; Kawai et al., 2004; Ning et al., 2011).

While IRF7 heavily influences the action of the interferon pathways, it is not itself inert to regulation. GBP4 is a regulatory factor that represses the function of IRF7 through direct inhibition of the phosphorylation process critical to the IRF7's function (Hu et al., 2011; Sun & Wang, 2012). GBP4 is a member of the guanylate binding protein family, a family of genes found both in mice and humans that has been shown to interact broadly with the innate immune response to a variety of threats including viruses (Braun et al., 2019; Selleck et al., 2013; Vestal & Jeyaratnam, 2011). The inhibition of IRF7 function by GBP4 could provide critical pressure to tip

the scales of immune response to HIV away from neurodegeneration in the brain. A visual summary of the relationship between the interferon response, IRF7 and GBP4 is presented in Figure 1.

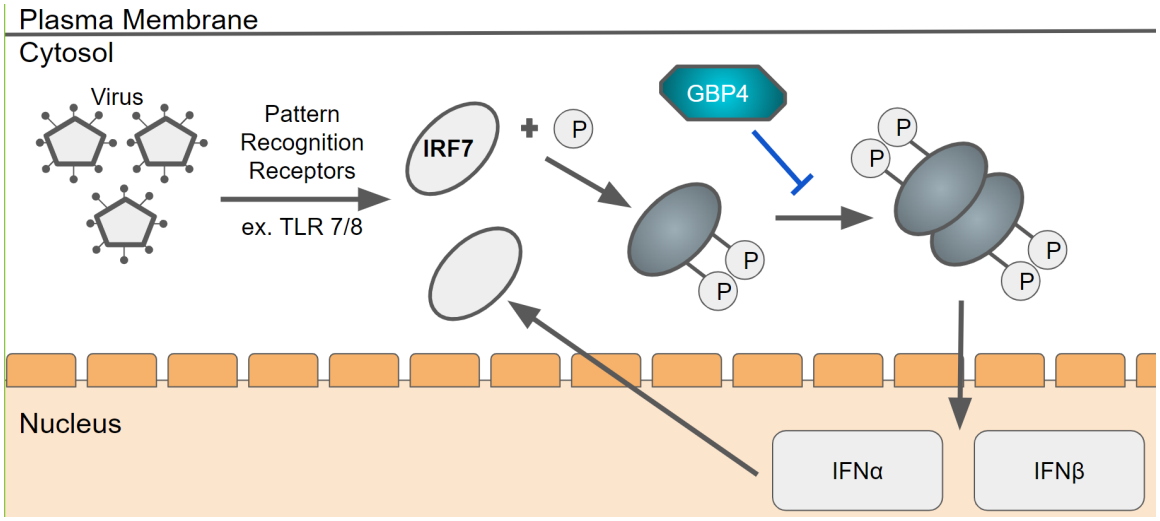


Fig 1. IRF7 and GBP4 in Type I Interferon Response to Virus. In a simplified representation, IRF7 is phosphorylated resulting from the activity of pattern recognition receptors after exposure to intracellular virus. This phosphorylation is required to enter the nucleus and stimulate genes in the type I and II interferon pathways. The type I interferon pathway then in turn produces more IRF7, which can be activated via intracellular virus. GBP4 is an interferon-stimulated gene in the guanylate binding protein family, which has been shown to prevent the phosphorylation and therefore activation of IRF7 during viral infection.

Genes in the guanylate binding protein gene family have demonstrated functions in controlling bacterial and viral infection, including reducing the infectivity of HIV (Braun et al., 2019; Degrandi et al., 2013; Ghimire et al., 2018; Krapp et al., 2016; Tretina et al., 2019). These genes are interferon-induced themselves, primarily by the type II IFN response (IFNɣ). However, GBP family members such as GBP1, GBP2 and GBP5 have been previously identified as upregulated in response to HIV viral infection (Vestal & Jeyaratnam, 2011).

Studying HIV-induced deleterious neurological patterns presents a distinct challenge. Research on living human tissue can only be conducted in separated human cells in culture, which sacrifices the biological context of a full organism with diverse interactions between and within organ systems. We are able to capture some of this diversity in communication using mouse models of NeuroHIV, but we must be cognizant of genetic and physiological differences between the model organism and the human health outcomes we wish to alter. Mouse GBPs have developed in an orthologous pattern compared to the gene family in humans (Olszewski et al., 2006). The mouse ortholog of GBP4, Gbp3 has been shown to be upregulated in mouse models of NeuroHIV [Kaul lab data, unpublished]. I then set out to determine methodologies for identifying GBP cellular contribution with tissue staining analysis. I report our findings here, including observation of distinct staining populations of immune cells in the brain. The first chapter of this dissertation will focus on this GBP-IRF7-neuroHIV relationship.

The relevance of specific genes within the context of HIV-induced neurodegeneration is often examined via cell quantification, as well as by gene expression analysis. This research relies on tissue imaging to extract information about disease progression. While advances in technology have allowed an explosion in available imaging techniques and ability to store and move digital images, quantification is still often done by hand. While quantification performed by a single researcher is held up as the gold standard for extracting information from images, this massively limits the speed at which critical research can be performed in ongoing health crises, such as that of HIV infection (Jensen, 2013).

Additionally, hand counts from individual researchers can be difficult to standardize (von Bartheld et al., 2016). Dividing work between researchers can increase the speed of overall

processing, but creates disparities between results from different researchers' counts, arising from innate biases. To avoid this issue, most cell quantification projects fall to a single researcher for consistency. However, even in this case, fatigue and increasing experience make it difficult to retain the same decision criteria or biases across an entire dataset which may constitute hundreds of images and thousands of cells. These difficulties have prompted the development of automatic cell counting methods which can address the issue of timeliness with mathematical consistency.

The increase in available processing power has also led to a meteoric rise in the prevalence of computerized methods to quantify cell information in images. Programs such as Unet use advanced neural networks run on large computing clusters, and while such programs are effective they are often out of reach for many research laboratories due to prohibitive barriers of cost and required expertise (Falk et al., 2019; Morelli et al., 2021). Other programs use machine learning to bring automated quantification to labs with greater accessibility, but they often do not include the full suite of tools to assess and validate model accuracy in the cell quantification task (Berg et al., 2019).

Therefore, the second chapter in this dissertation describes ACCT: Automatic Cell Counting with Trainable Weka Segmentation [Kataras et al. submitted for publication]. ACCT is built on the machine learning interface of Trainable Weka Segmentation (TWS) in ImageJ, a ubiquitous free image processing and analysis platform. We synthesize the powerful and intuitive training process of TWS and the scripting and processing capability within ImageJ and Python to produce a suite of code that guides the user through training, validation, full experimental analysis, and a final dataset audit to ensure the accuracy and consistency of cell counting. All of

this process is done without requiring the use of large external computing resources or any manipulation of code.

The modern research stage is defined by plentiful data, but limited by understanding. Our ability to address neurological dysfunction behind the blood-brain barrier is limited by the restriction of treatment penetrance and difficulty in studying this delicate closely-guarded system. This dissertation seeks to provide tools and guidance for researchers pushing forward the development of treatments for neuroHIV. In the process, I have identified Gbp4 as a gene that may shape the deleterious chronic immune activation during HIV infection. I have also produced and validated ACCT to enhance speed and consistency of cell quantification experiments with minimal barriers to accessibility.

References

Berg, S., Kutra, D., Kroeger, T., Straehle, C. N., Kausler, B. X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., Eren, K., Cervantes, J. I., Xu, B., Beuttenmueller, F., Wolny, A., Zhang, C., Koethe, U., Hamprecht, F. A., & Kreshuk, A. (2019). ilastik: interactive machine learning for (bio)image analysis. *Nature Methods*, *16*(12), 1226–1232.

Braun, E., Hotter, D., Koepke, L., Zech, F., Groß, R., Sparrer, K. M. J., Müller, J. A., Pfaller, C. K., Heusinger, E., Wombacher, R., Sutter, K., Dittmer, U., Winkler, M., Simmons, G., Jakobsen, M. R., Conzelmann, K.-K., Pöhlmann, S., Münch, J., Fackler, O. T., … Sauter, D. (2019). Guanylate-Binding Proteins 2 and 5 Exert Broad Antiviral Activity by Inhibiting Furin-Mediated Processing of Viral Envelope Proteins. *Cell Reports*, *27*(7), 2092–2104.e10.

D'Agostino, P. M., Gottfried-Blackmore, A., Anandasabapathy, N., & Bulloch, K. (2012). Brain dendritic cells: biology and pathology. *Acta Neuropathologica*, *124*(5), 599–614.

Degrandi, D., Kravets, E., Konermann, C., Beuter-Gunia, C., Klümpers, V., Lahme, S., Wischmann, E., Mausberg, A. K., Beer-Hammer, S., & Pfeffer, K. (2013). Murine guanylate binding protein 2 (mGBP2) controls Toxoplasma gondii replication. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(1), 294–299.

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., Dovzhenko, A., Tietz, O., Dal Bosco, C., Walsh, S., Saltukoglu, D., Tay, T. L., Prinz, M., Palme, K., Simons, M., … Ronneberger, O. (2019). U-Net: deep learning for cell counting, detection, and morphometry. *Nature Methods*, *16*(1), 67–70.

Ghimire, D., Rai, M., & Gaur, R. (2018). Novel host restriction factors implicated in HIV-1 replication. *The Journal of General Virology*, *99*(4), 435–446.

Gomez-Nicola, D., & Perry, V. H. (2015). Microglial dynamics and role in the healthy and diseased brain: a paradigm of functional plasticity. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *21*(2), 169–184.

Grant, I., Marcotte, T. D., & Heaton, R. K. (1999). Neurocognitive Complications of HIV Disease. *Psychological Science*, *10*(3), 191–195.

Honda, K., Yanai, H., Negishi, H., Asagiri, M., Sato, M., Mizutani, T., Shimada, N., Ohba, Y., Takaoka, A., Yoshida, N., & Taniguchi, T. (2005). IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature*, *434*(7034), 772–777.

Hu, Y., Wang, J., Yang, B., Zheng, N., Qin, M., Ji, Y., Lin, G., Tian, L., Wu, X., Wu, L., & Sun, B. (2011). Guanylate binding protein 4 negatively regulates virus-induced type I IFN and antiviral response by targeting IFN regulatory factor 7. *Journal of Immunology* , *187*(12), 6456–6462.

Jensen, E. C. (2013). Quantitative analysis of histological staining and fluorescence using ImageJ.

Anatomical Record , *296*(3), 378–381.

Kanmogne, G. D., Fonsah, J. Y., Umlauf, A., Moul, J., Doh, R. F., Kengne, A. M., Tang, B., Tagny, C. T., Nchindap, E., Kenmogne, L., Franklin, D., Njamnshi, D. M., Kuate, C. T., Mbanya, D., Njamnshi, A. K., & Heaton, R. K. (2020). Effects of HIV infection, antiretroviral therapy, and immune status on the speed of information processing and complex motor functions in adult Cameroonians. *Scientific Reports*, *10*(1), 14016.

Kaul, M., Garden, G. A., & Lipton, S. A. (2001). Pathways to neuronal injury and apoptosis in HIV-associated dementia. *Nature*, *410*(6831), 988–994.

Kawai, T., Sato, S., Ishii, K. J., Coban, C., Hemmi, H., Yamamoto, M., Terai, K., Matsuda, M., Inoue, J.-I., Uematsu, S., Takeuchi, O., & Akira, S. (2004). Interferon-alpha induction through Toll-like receptors involves a direct interaction of IRF7 with MyD88 and TRAF6. *Nature Immunology*, *5*(10), 1061–1068.

Krapp, C., Hotter, D., Gawanbacht, A., McLaren, P. J., Kluge, S. F., Stürzel, C. M., Mack, K., Reith, E., Engelhart, S., Ciuffi, A., Hornung, V., Sauter, D., Telenti, A., & Kirchhoff, F. (2016). Guanylate Binding Protein (GBP) 5 Is an Interferon-Inducible Inhibitor of HIV-1 Infectivity. *Cell Host & Microbe*, *19*(4), 504–514.

Laurence, J. (1993). T-cell subsets in health, infectious disease, and idiopathic CD4+ T lymphocytopenia. *Annals of Internal Medicine*, *119*(1), 55–62.

Morelli, R., Clissa, L., Amici, R., Cerri, M., Hitrec, T., Luppi, M., Rinaldi, L., Squarcio, F., & Zoccoli, A. (2021). Automating cell counting in fluorescent microscopy through deep learning with c-ResUnet. *Scientific Reports*, *11*(1), 22920.

Ning, S., Pagano, J. S., & Barber, G. N. (2011). IRF7: activation, regulation, modification and function. *Genes and Immunity*, *12*(6), 399–414.

Olszewski, M. A., Gray, J., & Vestal, D. J. (2006). In silico genomic analysis of the human and murine guanylate-binding protein (GBP) gene clusters. *Journal of Interferon & Cytokine Research: The Official Journal of the International Society for Interferon and Cytokine Research*, *26*(5), 328–352.

Saloner, R., & Cysique, L. A. (2017). HIV-Associated Neurocognitive Disorders: A Global Perspective. *Journal of the International Neuropsychological Society: JINS*, *23*(9-10), 860–869.

Savan, R. (2014). Post-transcriptional regulation of interferons and their signaling pathways. *Journal of Interferon & Cytokine Research: The Official Journal of the International Society for Interferon and Cytokine Research*, *34*(5), 318–329.

Selleck, E. M., Fentress, S. J., Beatty, W. L., Degrandi, D., Pfeffer, K., Virgin, H. W., 4th, Macmicking, J. D., & Sibley, L. D. (2013). Guanylate-binding protein 1 (Gbp1) contributes to cell-autonomous immunity against Toxoplasma gondii. *PLoS Pathogens*, *9*(4), e1003320.

Singh, H., Ojeda-Juárez, D., Maung, R., Shah, R., Roberts, A. J., & Kaul, M. (2020). A pivotal role for Interferon-α receptor-1 in neuronal injury induced by HIV-1. *Journal of Neuroinflammation*, *17*(1), 226.

Sun, B., & Wang, J. (2012). Guanylate binding protein (GBP) 4 negatively regulates virus induced type I interferon and antiviral response by targeting interferon regulatory factor (IRF) 7 (168.30). *The Journal of Immunology*, *188*(1 Supplement), 168.30–168.30.

Thaney, V. E., & Kaul, M. (2019). Type I Interferons in NeuroHIV. *Viral Immunology*, *32*(1), 7–14.

Thaney, V. E., O'Neill, A. M., Hoefer, M. M., Maung, R., Sanchez, A. B., & Kaul, M. (2017). IFNβ Protects Neurons from Damage in a Murine Model of HIV-1 Associated Brain Injury. *Scientific Reports*, *7*, 46514.

Tretina, K., Park, E.-S., Maminska, A., & MacMicking, J. D. (2019). Interferon-induced guanylate-binding proteins: Guardians of host defense in health and disease. *The Journal of Experimental Medicine*, *216*(3), 482–500.

Vestal, D. J., & Jeyaratnam, J. A. (2011). The guanylate-binding proteins: emerging insights into the biochemical properties and functions of this family of large interferon-induced guanosine triphosphatase. *Journal of Interferon & Cytokine Research*, 31(1), 89–97.

von Bartheld, C. S., Bahney, J., & Herculano-Houzel, S. (2016). The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *The Journal of Comparative Neurology*, 524(18), 3865–3895.

Examining GBP involvement in HIV induced neurodegeneration

Theodore J. Kataras[1,2], Hina Singh[2], Ricky Maung[2], and Marcus Kaul[1,2,*]

[1]University of California, Riverside, Graduate Program of Genetics, Genomics and Bioinformatics, 92507, USA
[2]University of California, Riverside, Dept. Biomedical Sciences, 92507, USA

Abstract:

Human Immunodeficiency Virus type 1 (HIV-1, hereafter referred to as HIV) infects $CD4^+$ immune cells, lymphocytes and macrophages as well as microglia in the brain. This causes chronic activation of the innate immune system which also enhances the activity and number of microglia, the resident immune cell of the brain. Chronic activation of microglia leads to the release of neurotoxic factors which cause neurological disorder as well as exacerbate aging related neurological diseases.

The guanylate binding protein family of genes (GBP) are Interferon (IFN)-induced cell-autonomous immune response genes to viruses and other pathogens. *GBP* genes have been tied to the IFN responses for specific pathogens, including HIV. We have found *Gbp3* to be upregulated in mouse brains expressing transgenic HIV envelope protein gp120 as transgene (gp120tg), as well as in HIV infected human macrophage cultures.

Additionally, in the gp120tg mouse model of HIV brain infection (NeuroHIV), *Gbp3* expression was correlated with imputed increases in microglia population size and HIV protein presence in the brain. *GBP4* (murine ortholog Gbp3) was one of the most significantly

upregulated genes in RNA-sequencing analysis of HIV infected human macrophages. We hope to further define the distinct pathways of GBP interface with host-pathogen interactions in HIV-induced neurodegeneration.

Introduction

HIV and the neuroimmune system

HIV was discovered in 1983 as the cause of Acquired Immunodeficiency Syndrome (AIDS) (Barré-Sinoussi, 1983; Gallo et al., 1984; Gallo and Montagnier, 2003; Montagnier, 2002, Popovic et al., 1984). In 2021 HIV affected 38.4 million people worldwide (UNAIDS, 2022). HIV primarily infects two types of immune cells: Activated CD4+ lymphocytes and macrophages. This infection causes a dramatic systemic drop in $CD4^+$ and $CD8^+$ lymphocyte numbers, eventually leading to AIDS (Laurence, 1993). Though the lymphocytes are far more numerous and suffer the largest decline, the macrophages are of particular interest due to their pivotal role in bringing the HIV infection into the central nervous system (CNS). The disorders and diseases associated with HIV infection of the CNS were originally called NeuroAIDS and are now generally referred to as NeuroHIV.

The pathology of NeuroHIV includes a chronic activation of macrophages and microglia in the brain. This sustained activation can further increase the activity and number of microglia as well as astrocytes in the brain (Borjabad et al., 2010; Kaul et al., 2001).

Microglia are the resident CNS macrophages, and are chronically activated by HIV infection, contact with HIV protein, or through cell signaling from nearby activated or injured cells. This activation can lead to neurotoxicity (Borjabad et al., 2010; Kaul et al., 2001). Both

circulating macrophages and microglia are immune cells with receptors for HIV binding and the capacity to release cytokines and promote apoptosis in response to a variety of pathogens (DePaula-Silva et al., 2019; Perry and Gordon, 1988). However, microglia differ from circulating macrophages in development and distinct response profiles (DePaula-Silva et al., 2019; Perry and Gordon, 1988). Astrocytes are glial cells that are critical for regulating and maintaining healthy synaptic functionality of neurons. They cannot sustain HIV infection, but have receptors for HIV protein binding and are activated by HIV proteins, which contribute to neurodegeneration. However, there is a gap in available knowledge directly linking the genetic mechanics of immune activation to the inflammatory cytokines and neurotoxins produced by activated microglia and astrocytes which cause neurodegeneration. This is evidenced by the lack of broadly effective treatments or biomarkers for disease progression in the NeuroHIV (Ellis et al., 2009; Kanmogne et al., 2020; Major et al., 2000).

The CNS injury caused by HIV results from inflammation, not only infection. HIV exposure causes neuronal damage through the activation of the IFN pathway leading to the release of cytotoxic factors, including cytokines, lipid mediators, reactive oxygen species (ROS) and excitotoxic amino acids. This leads to atrophy of neuronal synapses, dendrites and eventually neuronal death in rodent *in vitro* and *in vivo* models and human *in vitro* studies (Saylor et al., 2016). This HIV driven neuronal injury is also evidenced in human *ex-vivo* brain studies of HIV infected individuals (Kaul et al., 2001). Importantly, neuronal injury can be caused equally by recombinant viral protein without infection or by viral proteins released by infected macrophages and microglia (Ellis et al., 2009; Iskander et al., 2004).

HIV-induced neurodegeneration can be modeled with HIV gp120-transgenic (tg) mice, which constitutively express the soluble viral envelope protein gp120 in astrocytes in the mouse brain under the control of the promoter for glial fibrillary acidic protein (*GFAP*). HIVgp120tg mice recapitulate several of the hallmark neurodegenerative features observed in the brains of humans with NeuroHIV including: 1) decreased neuronal synaptic and dendritic density; 2) an increase in activated microglia and astrocytes; 3) patterns of differential gene expression; 4) behavioral deficits (D'Hooge et al., 1999; Krucker et al., 1998; Maung et al., 2014; Toggas et al., 1994).

The Guanylate Binding Protein Family

The guanylate binding protein (GBP) gene family has been implicated in pathogen-specific cell-autonomous immunity induced by IFN (Kim et al., 2012; MacMicking, 2012; Tripal et al., 2007).The GBP family is composed of 7 members in humans (hGBP) and 11 in mice (mGbp) with non-parallel homology (Olszewski et al., 2006). Specifically, Mouse Gbp1, Gbp2, Pseudogene Gbp1, Gbp3 and Gbp5 are respectively orthologous to human GBP1, GBP2, GBP3, GBP4 and GBP5.

Individual GBP proteins have been tied to HIV-specific immune responses: *GBP2* and *5* have been shown to reduce HIV infectivity in human immune cell culture by preventing viral envelope formation. The reduction in HIV infectivity from *GBP2* and *5* results from direct GBP binding to furin, a protease which cleaves proteins to activate their biological function, and is necessary for HIV envelope formation (Braun et al., 2019; Krapp et al., 2016).

Several other GBP family members have been found to be upregulated in response to specific HIV viral proteins (Braun et al., 2019; Fanibunda et al., 2013; Krapp et al., 2016;

Woollard et al., 2014). One of these upregulated genes, *Gbp4* inhibits the function of *Irf7*, an inducer and promoter of the type 1 IFN pathway, during Sendai virus infection of mouse macrophages, and has been implicated in human lung culture with viral infection as well. In human cells, *GBP4* competitive binding to *IRF7* is responsible for the repression of transcriptional regulation of IFN response without GTPase activity (Hu et al., 2011, p. 20; Kim et al., 2012; Sun and Wang, 2012, p. 4).

The IFN pathway is a well-studied immune response pathway which has two types, or paths of response. Type 1 IFNs (*IFN-α* and *IFN-β*) are generally involved in the inflammatory response to viral infection and other intracellular pathogens. Type 2 IFN (*IFN-γ*) responds to bacterial threats and fills an immunomodulatory role. GBPs as a whole are preferentially induced by the type 2 IFN, but are upregulated by HIV infection in macrophages and likely microglia as well (Appelberg et al., 2017). *IFN-β*, of type 1 IFN response, has been identified as critical to neuroprotection in the brain against HIV-induced inflammation (Barber et al., 2004; Thaney et al., 2017).

GBP4's murine ortholog was significantly upregulated in the HIVgp120tg mouse model, and the gene's expression was correlated with the estimated size of the microglial cell population. Mouse *Gbp4* inhibits the type 1 IFN pathway downstream of *IFN-α* or *IFN-β* receptor signaling *in vitro* (Sun and Wang, 2012, p. 4). The *IFN-β* derived neuroprotection was reliant on the presence of the *IFN-α* receptor, and altering the throughput of *IFN-α* and *IFN-β* could lead to a more neuroprotective antiviral state during HIV infection.

I also conducted an additional, ancillary study of Gbp3 localization though mouse brain tissue staining and imaging to expand on the findings from our PSEA analysis of the mouse model of neuroHIV. In determining adequate imaging methodologies through antibody titration and costaining analysis, I observed broad but variable distribution of GBP in Resting and Activated tissues as well as differential staining in populations of immune cells in the cortex partial to either Iba-1 or P2ry12 staining.

Materials and Methods:

Population Specific Enrichment Analysis

The brains of HIVgp120tg mice in addition to wild type mice of the same genetic background were used in microarray analysis to study the genetic factors involved in HIV-induced neurodegeneration (Maung et al., 2014; Toggas et al., 1994). In brief, the study includes only the wild type and HIVgp120tg animals from a cross involving a HIVgp120tg HIPEX strain founder line crossed with CCR5 knockout mice (CCR5KO, B6.129P2-Ccr5tm1Kuz/J) purchased from The Jackson Laboratory. The resultant F2 animals heterozygous for CCR5 and HIVtgp1201tg were used to generate the new mouse line (Maung et al., 2014).

In our Population Specific Enrichment Analysis (PSEA), we chose 4 of the most prevalent cell types in the brain for population estimation: neurons, astrocytes, microglia, and oligodendrocytes. We approximated each cell population size with a combination of probes from multiple genes tied to each cell type in the literature (citations and probe IDs provided in **Appendix A,** in brief neuron: Nefm, Gad1; astrocyte: Gfap, Aldh1l1, ; microglia: Cd68, Aif1; oligodendrocyte: Mog, Mbp, Mag). The genes were chosen based on primary literature and their

probes selected to maximize agreement within cell types while minimizing overlap with other cell types (**Fig 1.**). The quantitative cutoffs for probe inclusion in cell population estimates were total mean R^2 within cell populations > 0.5, and other probe sets < 0.5. This is similar to the methodology reported in (Capurro et al., 2014).

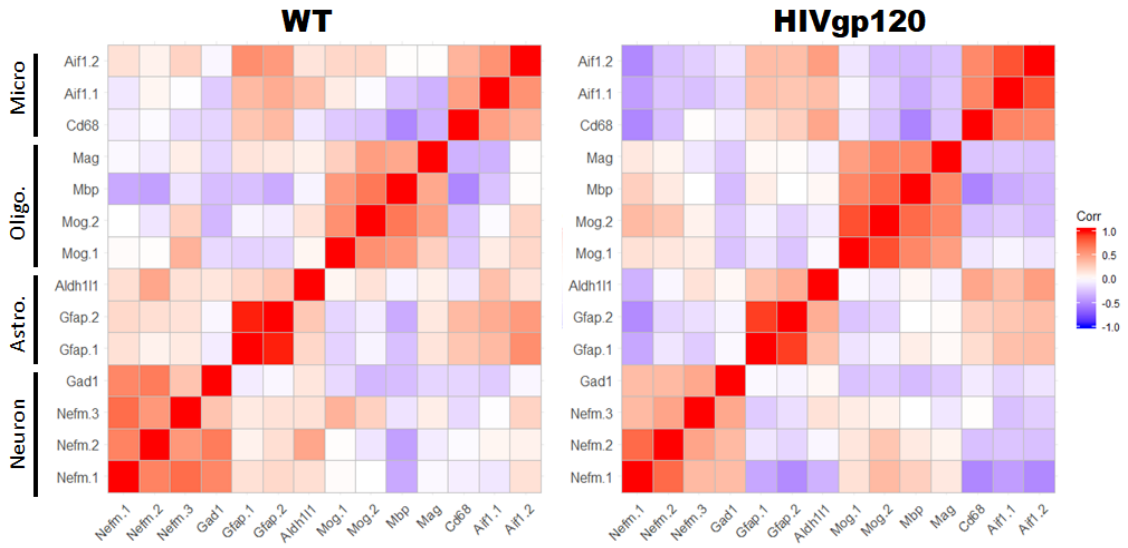


Fig 1. Correlation coefficient plot of wild type and HIVgp120tg used for cell population size estimation in PSEA. Selection criteria: total mean R^2 within cell populations > 0.5, and other probesets < 0.5. Multiple probes used for individual genes indicated by numbers after the period. Citations and probe IDs provided in **Appendix A**.

We then used these estimates to identify genes specifically associated with only a single cell type and upregulated in response to HIV via multiple linear regression. The linear models are constructed modularly from combinations of the 4 cell population estimates and the HIV protein presence regressor (**Fig. 2**). This methodology for identifying cell type specific pathogen response genes was previously applied to Huntington's Disease (Kuhn et al., 2015).

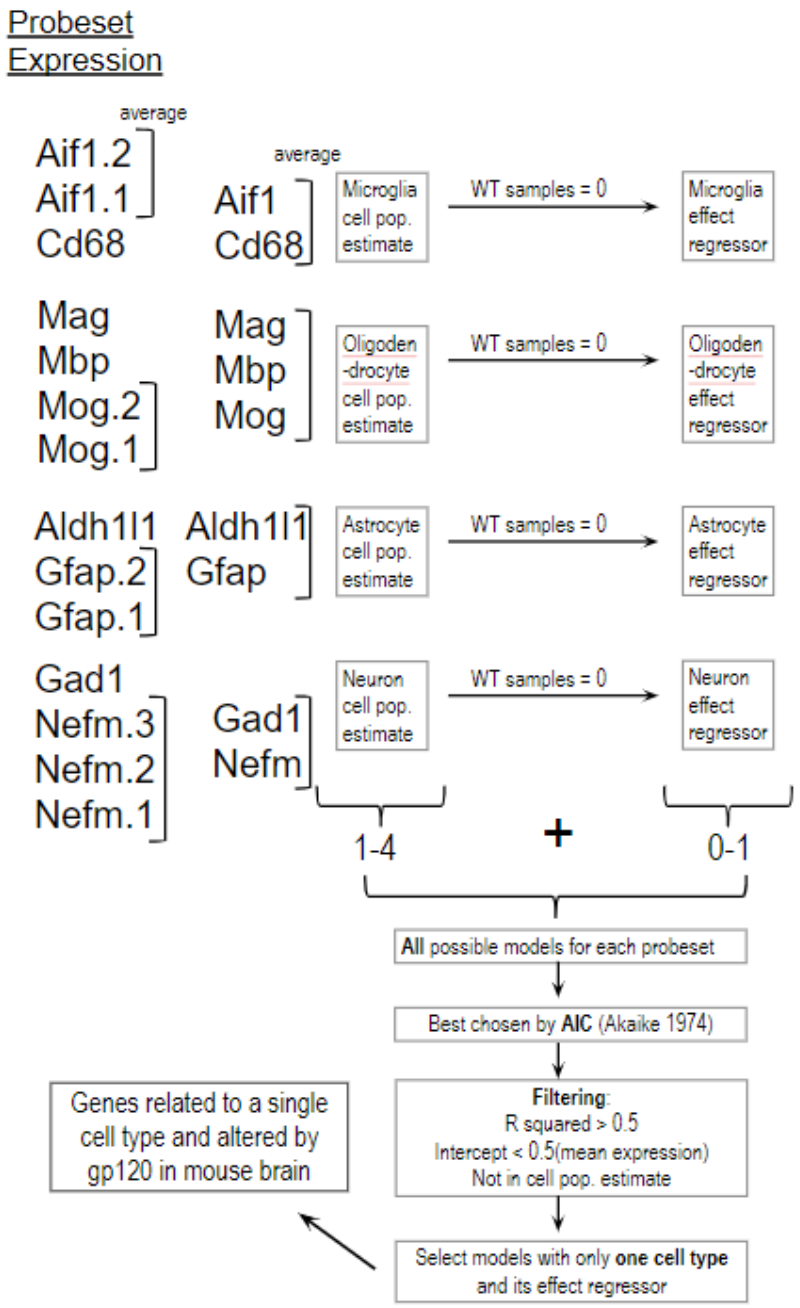


Fig. 2 Explanation of PSEA process on mouse brain microarray data. Single cell type expression was assigned by computing the correlation between each probeset and the reference signals and restricting further analyses to the cases where the correlation was larger than 0.8 for a single cell type (e.g., neurons) and less than 0.2 for the other three cells types (e.g., astrocyte, oligodendrocyte and microglia). We then calculated regressions with only one cell-type regressor (neurons in the above example) plus the corresponding auxiliary regressor (for differential expression between disease and control). This general methodology for estimating changing cell populations in mixed tissue analysis was designed by Alexander Kuhn in 2015.

Monocyte Derived Macrophage RNA Sequencing

Human monocyte-derived macrophage (MDM) cell culture: white blood cell concentrates (WBC) were purchased from the San Diego Blood Bank and further purified via a Ficoll-Paque density gradient centrifugation as described in previous work by the Kaul lab (Medders et al., 2010).

The fraction containing monocytes, which can be differentiated into macrophages, were isolated. islation was achived by centrifugation of heparinized blood at 200g for 20 min, buffy coat cells using density gradients of Ficoll-paque (r 1.073; GE Health Life Sciences, Piscataway, NJ) and transferred into 75-cm2 cell-culture flasks. Media with human serum was then used for *in vitro* differentiation into macrophages. after allowing adherence in RPMI 1640 containing 2 mM glutamine, antibiotics (penicillin/streptomycin), and 10% human AB serum (RPMI-ABS) at 37℃, 6% CO_2, in humidified atmosphere. Nonadherent cells were removed by rigorous washing with warm RPMI 1640 (37℃). Adherent monocytes were then cultured for 7 days in the above serum-containing medium to allow for differentiation into MDMs. Cells were detached and harvested for experimentation with a rubber policeman in a rinse of PBS containing Ethylenediaminetetraacetic acid (EDTA) (0.2 g/l; Sigma, St. Louis, MO) for 5–10 min at 37℃.

Cultures were paired after splitting, with two uninfected and two infected cultures per individual. Each culture was seeded with $5x10^5$ MDMs into 12 or 24 well plates after washing with PBS and cultured for 11-13 days in RPMI-ABS. Dead cells were identified with trypan blue to observe culture health. HIV infection was performed with type 1 HIV-BaL strain culture fluid acquired from the NIH and confirmed with detection of the p24 viral protein. Inoculation period

was 4 hours with 500μl of HIV-BaL followed by 3 washes in phosphate buffered saline. The cells were infected with HIV for 5 days before harvesting for analysis. RNA extraction was performed using the QIAshredder column and RNeasy minikit by Dr. Ana Sanchez [Sanchez et al., Kaul Lab unpublished]. Sequencing analysis was performed via sequencing with an Illumina HiSeq 2500 using both single and paired end reads. The raw data of the RNA-seq were then transferred to Cory White and Christopher Woelk at University of California, San Diego for analysis. Comparative expression values were calculated in EdgeR and normalized via the weighted trimmed mean of M-values methodology.

THP-1 macrophage RT-qPCR

To examine the potential for use in further studies on GBP4-HIV interaction, human monocytic THP-1 cells (macrophage-like) were cultured and interrogated with RT-qPCR analysis for RNA expression after stimulation with bacterial lipopolysaccharides (LPS) as an immune-stimulant.

Cells were grown in medium containing 90% RPMI 1640 (Life Technologies, Carlsbad, CA), 10% FBS (Hyclone), 0.05µM 2- mercaptoethanol, 2 µML-glutamine (Sigma), and a combination of 100 µml penicillin with 100 mg/ml streptomycin (Sigma) incubated at 37℃ with 5% CO2. The cells were split 1:3–1:4 once or twice a week after reaching $1x10^6$ cells/ml. The cells were exposed in a 6 well plate to LPS stimulation or PBS control for either 6 or 24 hours. Gene expression was evaluated for GBP1, GBP4, IRF7 and IFIT1 measured via purified RNA extract from each culture. RT-qPCR was performed using novel primers for GBP4 and IRF7, as well as previously validated primers for GBP1 and IFIT1. Novel primers were designed using Primer3 and Primer-Blast software, and reverse transcription was performed with 2-4µg of RNA.

Amplification reactions were performed using a QuantStudio 6 plus Flex Real-Time PCR system. RNA isolation, cDNA synthesis and PCR procedures were carried out as described in previous publication (Singh et al., 2020).

Mouse cortex staining

I additionally conducted preliminary analyses to determine adequate methodology for confirming Gbp3 localization in mouse brain via tissue imaging, specifically the cortex. I have included a description of the methodology and preliminary observations here as a baseline from which to begin further studies. Microscopy was performed with a Zeiss 200 M fluorescence deconvolution microscope with computer-controlled 3D stage and filters for DAPI, FITC, Cy3 and Cy5. All images were collected using Slidebook software (version 6, Intelligent Imaging Innovations, Inc., Denver, CO).This imaging analysis was performed using mouse 40µm thick frozen brain samples previously stored in cyropreservative medium (30% glycerol, 30%ethylene glycol in phosphate buffered saline (PBS). Tissue was selected from two male mice. Tissue was permeabilized with 0.5% Triton X-100 for 30 minutes and blocked with 10% goat serum for 1.5 hours. The mice had a mixed C57BL/6.129/SJL genetic background and were 5 months old at harvest. Mice were selected to include one wild type control (Resting, n = 1) and one transgenic (HIVgp120tg, Activated, n = 1). No randomization was performed.

For resolving Gbp3 localization in tissue, I tested concentrations of 8,13, and 18 µg/ml of Gbp3 antibody (Abbexa, abx176742) with Alexa 488 (Invitrogen, A11034) secondary antibody at dilution of 1:200 in PBST and Hoechst 33342 (ThermoFisher, 62249) nuclear staining at dilution of 1:150 in PBS. Primary staining was performed overnight, and secondary staining was performed for 1 hour at room temperature and nuclear staining for 5 minutes. A total of three 5

minute washes were performed between all permeabilizing, blocking and staining steps in PBS or PBST (PBS + 0.2% Tween-20).

Our chosen antibodies for Gbp3 and our preferred microglial marker, Iba-1, required the same secondary antibody hosting, so we next compared Iba-1 staining to staining for P2ry12, an additional microglial marker stain (Okunuki et al., 2018). P2ry12 antibody (BioLegend ,848001) was used at 1.25, 2.5, and 5 µg/ml for evaluation with Alexa 647 (Molecular Probes, A-31573) secondary antibody at 2.5 µg/ml. Iba-1 (Wako Chemicals , ID 019-19741) was used for staining at a dilution of 1:500 and visualized with Alexa 488 secondary antibody at dilution of 1:400.

We also examined tomato lectin staining in order to find a suitable marker for colocalization studies that would allow us to measure the changes in Gbp3 staining intensity in tissue locations specifically overlapping with microglia (Villacampa et al., 2013). We examined the use of a tomato lectin for visualizing the microglia along with Iba-1 staining for confirmation and Gbp3 staining as well. This lectin is known to stain immune cells in the brain as well as endothelial tissue such as blood vessels. We used Iba-1 and Gbp3 staining separately (as they both are rabbit IgG and rely on the same secondary antibody host) to compare with the immune cells visualized by Tomato lectin (Lycopersicon esculentum, Vector Laboratories Fl-1171) in wild type and transgenic animals (Fig. 5). Iba-1 and Gbp3 staining was run as described previously with Gbp3 antibody used at 2.5 µg/ml and secondary antibody Alexa 594 (Molecular Probes, A11037) for both at 1:400 dilution. The lectin was left on tissue for one hour during staining at dilution of 1:200. Hoechst nuclear staining was also performed as previously stated.

Results

The gp120 transgenic mouse brains displayed increased GBP expression and indicated microglia association of *GBP4* through Population Specific Enrichment Analysis (PSEA) (**Fig. 3a**). To extend our findings beyond the transgenic mouse brain microarray and into human cells, we examined the transcription of human monocyte-derived macrophage (MDM) culture with RNA Seq, and found that GBP genes were again upregulated by HIV infection in a similar pattern within the GBP family (**Fig. 3b**).

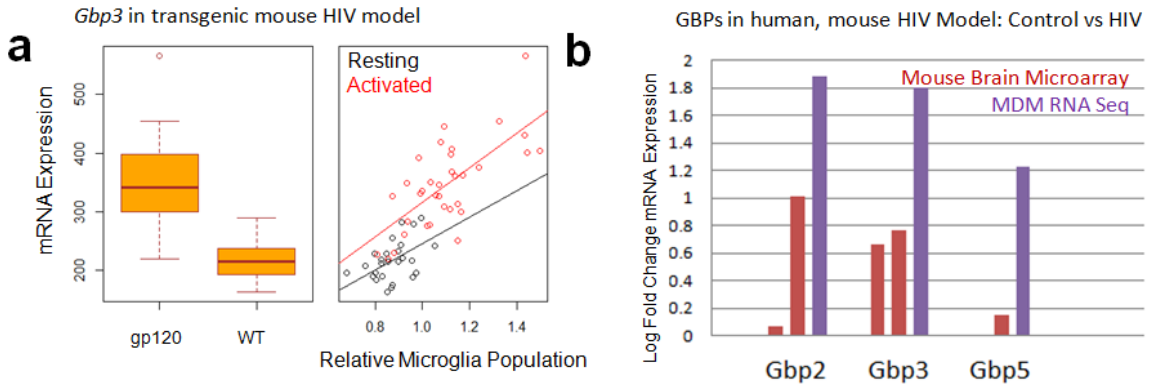


Fig 3. Guanylate Binding Proteins respond to HIV in immune cells inside and outside the brain. Microarray expression of HIVgp120tg mice associated murine *Gbp3* with microglia population and transgenic HIV protein presence via multiple linear regression, quantile plot t-test $P = 8.17e^{-12}$ and multiple regression with $R^2 = 0.74$, $P=2.2e^{-16}$ (**a**). Probes associated GBP genes and human orthologs in a previously described microarray (red) and an RNA Seq experiment on Human cultured Immune cells with HIV infection (purple) for 24 hours showed upregulation of GBP genes in HIV context (**b**).

These cells indicated a possible restrictive relationship between GBP4 and IRF7, where cultures with lower amounts of GBP4 showed a trend for greater amounts of IRF7, indicating that GBP4 may be restricting the self-amplification of virus-induced type 1 IFN signaling (**Fig. 4a**).

Additionally, we assessed a THP-1 macrophage culture with LPS stimulation with RT-qPCR to examine the validity of the model system and newly designed primers for human *GBP4* and *IRF7*. We found increased expression of *GBP4* at 24 hours of LPS stimulation vs 6 hours, and IRF7 was upregulated similarly at both of the time points (**Fig. 4b**). However, both increases were smaller in magnitude than the expression changes in GBP1 and IFIT1, which were increased in both time points. As this was a preliminary study, there was a single technical replicate for each condition and no biological replicates, so the results are to be taken as preliminary findings, without calculated statistical significance.

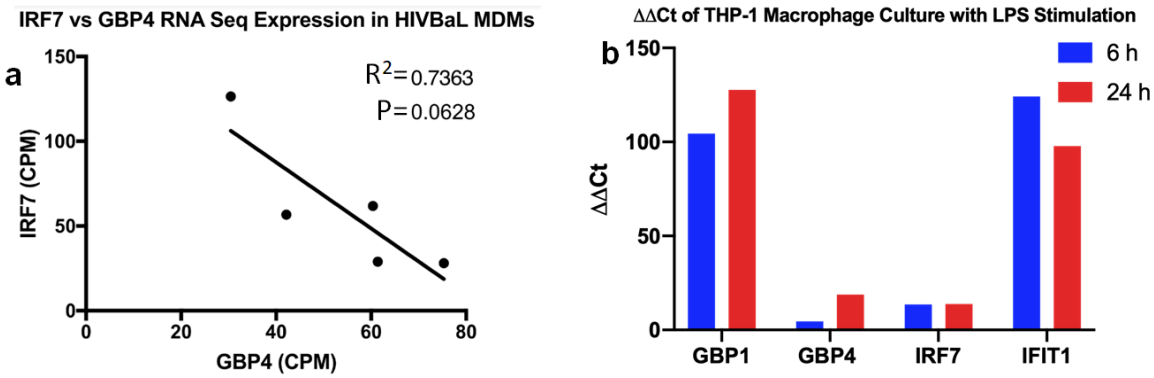


Fig 4. GBP4 and IRF7 in human macrophage culture. MDM from healthy donors with 6 day *in vitro* HIV infection, 2 technical replicates each. Expression calculated as counts-per-million reads (cpm) in EdgeR from RNA Sequencing on Illumina Hi Seq 2500. *GBP4* shows a negative trend with *IRF7* (a). A preliminary experiment with THP-1 human macrophage culture was used to validate primers at 6 and 24 hour time points with RT-qPCR. Of the ISGs *GBP4* showed less responsiveness to the LPS used to stimulate an immune response than *GBP1*, and *IFIT1*. IRF7 did not trend with GBP4 under these time points and stimulus, possibly indicating that HIV and LPS differ in the ways they engage the IFN response (b).

In our examination of tissue staining methods for further examining Gbp3 localization, we found similar definition of Gbp3 protein in cells at all tested antibody concentrations. This staining revealed the presence of Gbp3 in many cells in the cortex, not just microglia. However,

in the transgenic, Activated tissue images we saw heterogeneous cell shapes within the cortex that were distinct from other round cells (Fig. 5).

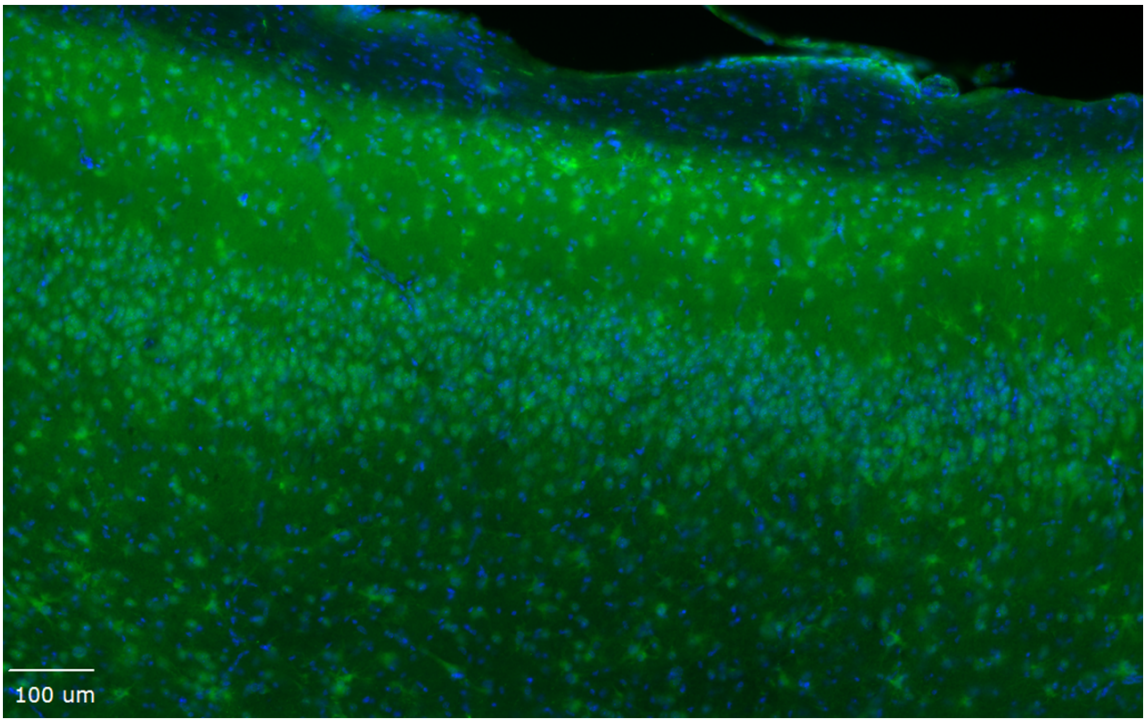


Fig. 5 Gbp3 antibody in Activated mouse cortex at 10x magnification. HIVgp120 cortex tissue stained at concentrations of 18 µg/ml of Gbp3 antibody with Alexa 484 secondary antibody (green) and Hoechst nuclear staining (blue).

These heterogeneously shaped cell objects were not visible in the wild type, Resting tissue, indicating that they have some connection to the innate immune response in our mouse model of HIV induced neurodegeneration (Fig. 6). Additionally, it was apparent from this initial imaging, this cell population seemed to appear in multiple layers throughout the cortex.

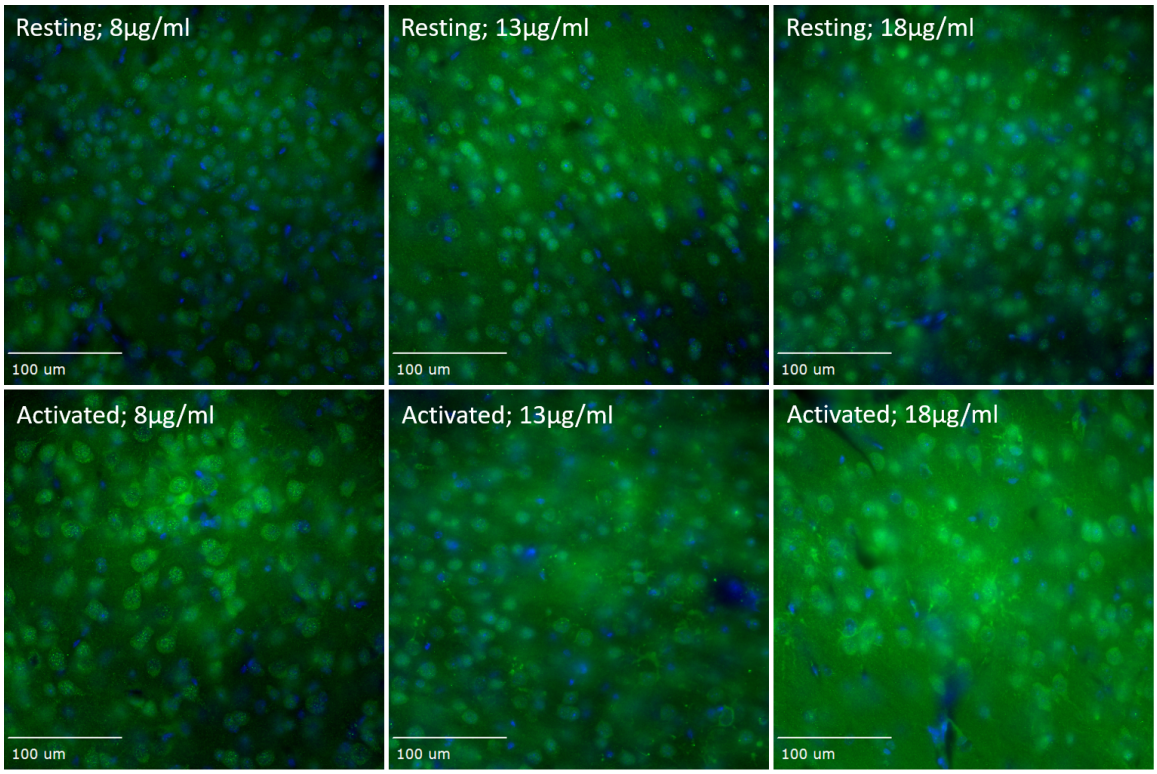


Fig. 6 Gbp3 antibody titration in mouse cortex. 40x magnification images of wild type, Resting (top row) and Activated, HIVgp120 (bottom row) tissue tested at Gbp3 antibody concentrations of 8,13, and 18 µg/ml with Alexa 488 secondary antibody (green) and Hoechst nuclear staining (blue).

The P2ry12 co-staining with Iba-1 revealed two separate, non-overlapping populations of stained cells. Iba-1 had a greater tendency to bind to processes than P2ry12 did, but I did not observe any overlapping cell body staining of the two antibodies (Fig. 7). This finding presents an avenue for further research into characterizing these two apparently distinct populations of cells, but given the disparity we could not use P2ry12 as a reliable marker to visualize microglia in the brain as we had with Iba-1 for comparison to Gbp3 localization.

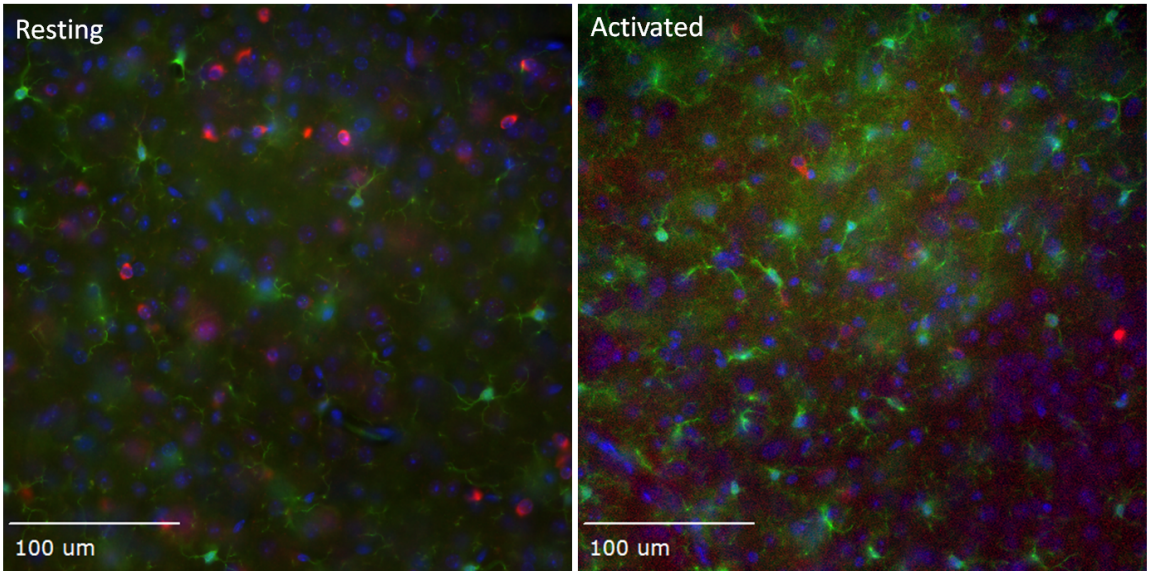


Fig. 7 Iba-1 and P2ry12 co-staining in mouse cortex. 40x magnification images of P2ry12 antibody (red) staining with Alexa 647 secondary antibody and Iba-1 antibody (green) with Alexa 488 and Hoechst nuclear staining (blue) of wild type (left) and HIVgp120tg tissue (right).

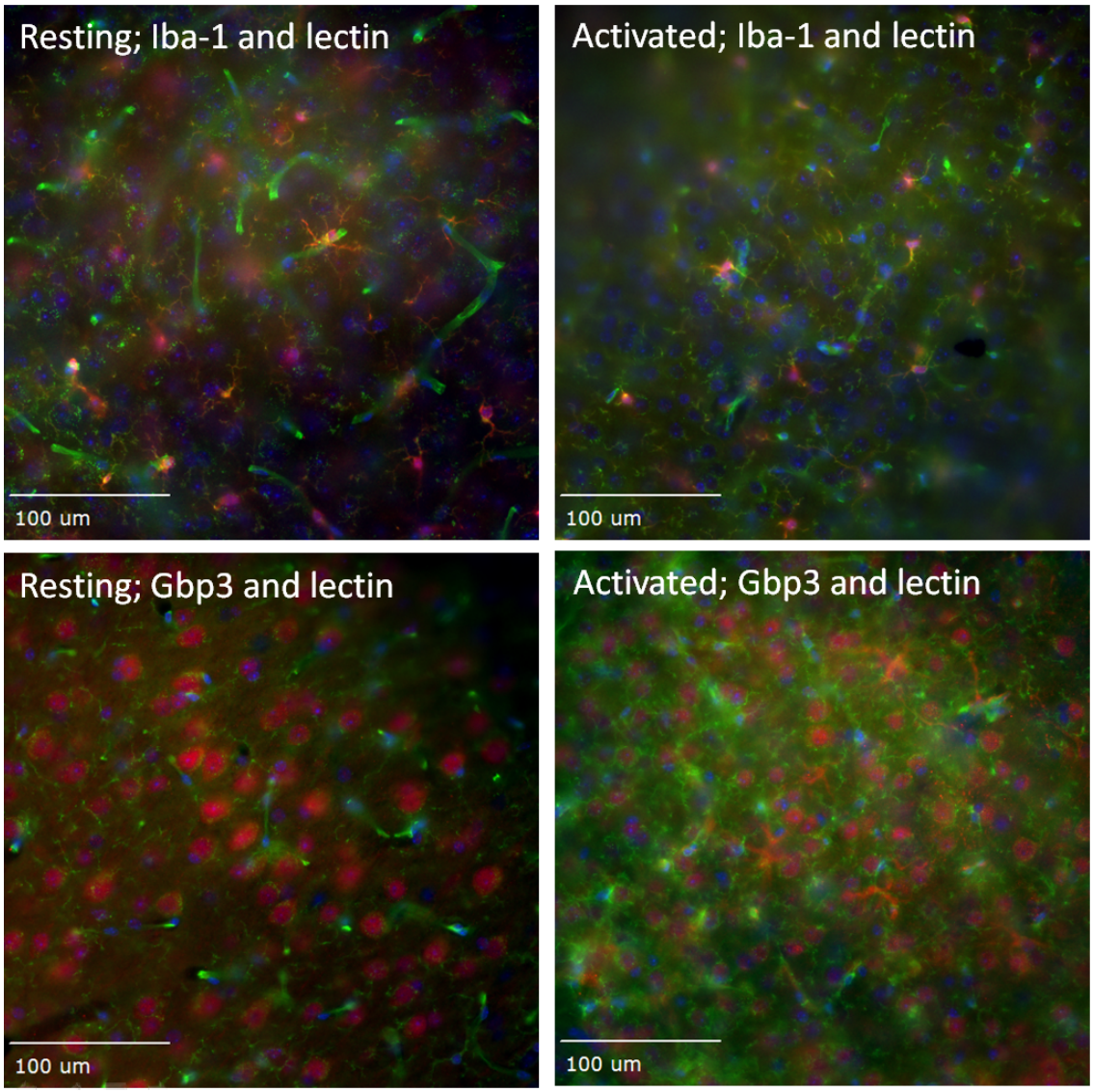


Fig. 8 Tomato lectin with Iba-1 or Gbp3 staining in mouse cortex. 40x magnification images of tomato lectin (Lycopersicon esculentum) staining (green) and Hoechst nuclear staining (blue) were used with Iba-1 (red, top row) or Gbp3 (red, bottom row) staining with Alexa 594 secondary antibody in wild type, Resting (left column) and HIVgp120tg, Activated (right column) mouse cortex tissue.

Last, I compared Gbp3 and Iba-1 staining to a tomato lectin staining (Fig. 8). Again, the Gbp3 staining in Activated tissue revealed heterogeneously shaped cell objects, but these did not colocalize with the lectin staining. The lectin, however, also showed little to no colocalization

with the Iba-1 staining for microglia. So, an alternative is still needed to confirm the identity of these cells and allow us to use image analysis techniques to measure specific Gbp3 expression localization. Despite the lack of antibodies for Gbp3 and Iba1 that are suitable for combined staining, the large number of cells labeled for Gbp3 clearly indicates that the protein is widely expressed in cerebral cortex and unlikely to be limited to microglia.

Discussion

GBPs are strongly induced by IFN signaling but for most members of the GBP family, the downstream effects of the proteins during immune cell activation are unclear. Our data and the related literature suggest that multiple members of the GBP family play roles in the response to HIV (DePaula-Silva et al., 2019; Hu et al., 2011; Krapp et al., 2016; Vestal and Jeyaratnam, 2011). We have identified a possible connection in the HIVgp120-transgenic mouse brain for Gbp3 (GBP4 mouse ortholog) and putative microglia population size due to HIV protein exposure. Additionally, I have identified a negative trend in expression of GBP4 and IRF7 in the HIV-infected human monocyte-derived macrophage culture. This interaction sheds light into a corner of the complex interplay between neuroprotection and neurodegeneration that takes place in the brain during HIV infection.

Preliminary mRNA analysis of THP-1 macrophages after stimulation with LPS indicates that GBP changes may be reduced in magnitude in this cell model. However, by comparison to uninfected control cultures, even the smaller changes are detectable via RT-qPCR, as demonstrated with the THP-1 cells and *GBP4*. These somewhat minute changes were not entirely unexpected, as the GBP proteins in general show decreased activation in iPSC-derived microglia in response to a bacterial lipopolysaccharide (LPS), a lipid-bound sugar, compared to *ex vivo* measures (Hasselmann et al., 2019).

However, GBP effects in circulating and brain resident immune cells in the context of HIV-induced neurodegeneration remain to be fully characterized. It will be important to determine which genes in the IFN response pathway may have regulatory relationships with GBP family members. All GBPs are IFN stimulated, but mGbp4/hu*GBP4* and possibly others interact with the IFN signaling itself within the cell, potentially acting as positive or negative feedback mechanisms.

Efforts to develop a methodology for Gbp3 colocalization with microglia revealed that the spread of Gbp3 was broader than previously anticipated, but that the differential expression may still be driven by the heterogeneous cell shape population that we have not yet been able to colocalize with accepted markers. Additionally, the tissue staining revealed two distinct and non-overlapping populations of immune cells in the cortex in both Resting and Activated tissues. The Iba-1 marker and P2ry12 marker for immune cell staining in the brain were found to isolate themselves to wholly separate cell bodies, which lays important groundwork for further characterizing these antibodies and the populations of cells they bind to.

Examination of the protein localization will help to further characterize the role of GBP4 in HIV response at the protein level, but our present analyses lay the groundwork for understanding the interaction between this particular member of the gene family and the interferon signaling system through IRF7.

Additional information

Correspondence should be addressed to M.K. and requests for materials/tools should be addressed to T.K., or M.K.

Competing interests

The authors declare no competing interests.

References

Adam SA, Schnell O, Pöschl J, et al. ALDH1A1 Is a Marker of Astrocytic Differentiation during Brain Development and Correlates with Better Survival in Glioblastoma Patients: ALDH1A1 in Brain Development and Glioblastomas. Brain Pathol 2012;22(6):788–797; doi: 10.1111/j.1750-3639.2012.00592.x.

Appelberg KS, Wallet MA, Taylor JP, et al. HIV-1 Infection Primes Macrophages Through STAT Signaling to Promote Enhanced Inflammation and Viral Replication. AIDS Res Hum Retroviruses 2017;33(7):690–702; doi: 10.1089/AID.2016.0273.

Barbarese E. Spatial Distribution of Myelin Basic Protein MRNA and Polypeptide in Quaking Oligodendrocytes in Culture. J Neurosci Res 1991;29(3):271–281; doi: 10.1002/jnr.490290302.

Barber, S. A., Herbst, D. S., Bullock, B. T., Gama, L., & Clements, J. E. (2004). Innate immune responses and control of acute simian immunodeficiency virus replication in the central nervous system. *Journal of Neurovirology*, *10 Suppl 1*, 15–20.

Barré-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., & Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, *220*(4599), 868–871.

Borjabad A, Brooks AI and Volsky DJ. Gene Expression Profiles of HIV-1-Infected Glia and Brain: Toward Better Understanding of the Role of Astrocytes in HIV-1-Associated Neurocognitive Disorders. J Neuroimmune Pharmacol 2010;5(1):44–62; doi: 10.1007/s11481-009-9167-1.

Braun E, Hotter D, Koepke L, et al. Guanylate-Binding Proteins 2 and 5 Exert Broad Antiviral Activity by Inhibiting Furin-Mediated Processing of Viral Envelope Proteins. Cell Rep 2019;27(7):2092-2104.e10; doi: 10.1016/j.celrep.2019.04.063.

Capurro A, Bodea L-G, Schaefer P, et al. Computational Deconvolution of Genome Wide Expression Data from Parkinson's and Huntington's Disease Brain Tissues Using Population-Specific Expression Analysis. Front Neurosci 2014;8:441; doi: 10.3389/fnins.2014.00441.

Deininger MH, Meyermann R and Schluesener HJ. The Allograft Inflammatory Factor-1 Family of Proteins. FEBS Lett 2002;514(2–3):115–121; doi: 10.1016/s0014-5793(02)02430-4.

DePaula-Silva AB, Gorbea C, Doty DJ, et al. Differential Transcriptional Profiles Identify Microglial- and Macrophage-Specific Gene Markers Expressed during Virus-Induced Neuroinflammation. J Neuroinflammation 2019;16(1):152; doi: 10.1186/s12974-019-1545-x.

D'Hooge R, Franck F, Mucke L, et al. Age-Related Behavioural Deficits in Transgenic Mice Expressing the HIV-1 Coat Protein Gp120: Behavioural Deficits in Gp120-Expressing Mice. Eur J Neurosci 1999;11(12):4398–4402; doi: 10.1046/j.1460-9568.1999.00857.x.

Ellis RJ, Calero P and Stockin MD. HIV Infection and the Central Nervous System: A Primer. Neuropsychol Rev 2009;19(2):144–151; doi: 10.1007/s11065-009-9094-1.

Fanibunda SE, Modi DN and Bandivdekar AH. HIV Gp120 Induced Gene Expression Signatures in Vaginal Epithelial Cells. Microbes Infect 2013;15(12):806–815; doi: 10.1016/j.micinf.2013.07.003.

Gallo RC and Montagnier L. The Discovery of HIV as the Cause of AIDS. N Engl J Med 2003;349(24):2283–2285; doi: 10.1056/NEJMp038194.

Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., & Safai, B. (1984). Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science*, *224*(4648), 500–503.

Hasselmann J, Coburn MA, England W, et al. Development of a Chimeric Model to Study and Manipulate Human Microglia In Vivo. Neuron 2019;103(6):1016-1033.e10; doi: 10.1016/j.neuron.2019.07.002.

Hu Y, Wang J, Yang B, et al. Guanylate Binding Protein 4 Negatively Regulates Virus-Induced Type I IFN and Antiviral Response by Targeting IFN Regulatory Factor 7. J Immunol 2011;187(12):6456–6462; doi: 10.4049/jimmunol.1003691.

Iskander S, Walsh KA and Hammond RR. Human CNS Cultures Exposed to HIV-1 Gp120 Reproduce Dendritic Injuries of HIV-1-Associated Dementia. J Neuroinflammation 2004;1(1):7; doi: 10.1186/1742-2094-1-7.

Kanmogne GD, Fonsah JY, Umlauf A, et al. Effects of HIV Infection, Antiretroviral Therapy, and Immune Status on the Speed of Information Processing and Complex Motor Functions in Adult Cameroonians. Sci Rep 2020;10(1):14016; doi: 10.1038/s41598-020-70981-4.

Kaul M, Garden GA and Lipton SA. Pathways to Neuronal Injury and Apoptosis in HIV-Associated Dementia. Nature 2001;410(6831):988–994; doi: 10.1038/35073667.

Kim B-H, Shenoy AR, Kumar P, et al. IFN-Inducible GTPases in Host Cell Defense. Cell Host Microbe 2012;12(4):432–444; doi: 10.1016/j.chom.2012.09.007.

Kodama T, Guerrero S, Shin M, et al. Neuronal Classification and Marker Gene Identification via Single-Cell Expression Profiling of Brainstem Vestibular Neurons Subserving Cerebellar Learning. J Neurosci 2012;32(23):7819–7831; doi: 10.1523/JNEUROSCI.0543-12.2012.

Krapp C, Hotter D, Gawanbacht A, et al. Guanylate Binding Protein (GBP) 5 Is an Interferon-Inducible Inhibitor of HIV-1 Infectivity. Cell Host Microbe 2016;19(4):504–514; doi: 10.1016/j.chom.2016.02.019.

Krucker T, Toggas SM, Mucke L, et al. Transgenic Mice with Cerebral Expression of Human Immunodeficiency Virus Type-1 Coat Protein Gp120 Show Divergent Changes in Short- and Long-Term Potentiation in CA1 Hippocampus. Neuroscience 1998;83(3):691–700; doi: 10.1016/s0306-4522(97)00413-2.

Kuhn A, Capurro A and Luthi-Carter R. Gene Expression-Based Approaches to Understanding Huntington's Disease and New Tools for the Interpretation of Expression Datasets. In: Applied Neurogenomics. (Jain KK. ed) Springer New York: New York, NY; 2015; pp. 61–91; doi: 10.1007/978-1-4939-2247-5_2.

Kuhn A, Thu D, Waldvogel HJ, et al. Population-Specific Expression Analysis (PSEA) Reveals Molecular Changes in Diseased Brain. Nat Methods 2011;8(11):945–947; doi: 10.1038/nmeth.1710.

Laurence J. T-Cell Subsets in Health, Infectious Disease, and Idiopathic CD4+ T Lymphocytopenia. Ann Intern Med 1993;119(1):55–62; doi: 10.7326/0003-4819-119-1-199307010-00010.

MacMicking JD. Interferon-Inducible Effector Mechanisms in Cell-Autonomous Immunity. Nat Rev Immunol 2012;12(5):367–382; doi: 10.1038/nri3210.

Major EO, Rausch D, Marra C, et al. HIV-Associated Dementia. Science 2000;288(5465):440–442; doi: 10.1126/science.288.5465.439d.

Maung R, Hoefer MM, Sanchez AB, et al. CCR5 Knockout Prevents Neuronal Injury and Behavioral Impairment Induced in a Transgenic Mouse Model by a CXCR4-Using HIV-1 Glycoprotein 120. J Immunol 2014;193(4):1895–1910; doi: 10.4049/jimmunol.1302915.

Medders KE, Sejbuk NE, Maung R, et al. Activation of P38 MAPK Is Required in Monocytic and Neuronal Cells for HIV Glycoprotein 120-Induced Neurotoxicity. J Immunol 2010;185(8):4883–4895; doi: 10.4049/jimmunol.0902535.

Montagnier L. Historical Essay. A History of HIV Discovery. Science 2002;298(5599):1727–1728; doi: 10.1126/science.1079027.

Okunuki, Y., Mukai, R., Pearsall, E. A., Klokman, G., Husain, D., Park, D.-H., Korobkina, E., Weiner, H. L., Butovsky, O., Ksander, B. R., Miller, J. W., & Connor, K. M. (2018). Microglia inhibit photoreceptor cell death and regulate immune cell infiltration in response to retinal detachment. Proceedings of the National Academy of Sciences of the United States of America. https://doi.org/10.1073/pnas.1719601115

Olszewski MA, Gray J and Vestal DJ. In Silico Genomic Analysis of the Human and Murine Guanylate-Binding Protein (GBP) Gene Clusters. J Interferon Cytokine Res 2006;26(5):328–352; doi: 10.1089/jir.2006.26.328.

Perry VH and Gordon S. Macrophages and Microglia in the Nervous System. Trends Neurosci 1988;11(6):273–277; doi: 10.1016/0166-2236(88)90110-5.

Popovic, M., Sarngadharan, M. G., Read, E., & Gallo, R. C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science*, *224*(4648), 497–500.

Quarles RH. Myelin-Associated Glycoprotein in Development and Disease. Dev Neurosci 1983;6(6):285–303; doi: 10.1159/000112356.

Saylor D, Dickens AM, Sacktor N, et al. HIV-Associated Neurocognitive Disorder--Pathogenesis and Prospects for Treatment. Nat Rev Neurol 2016;12(4):234–248; doi: 10.1038/nrneurol.2016.27.

Scolding NJ, Frith S, Linington C, et al. Myelin-Oligodendrocyte Glycoprotein (MOG) Is a Surface Marker of Oligodendrocyte Maturation. J Neuroimmunol 1989;22(3):169–176; doi: 10.1016/0165-5728(89)90014-3.

Singh H, Ojeda-Juárez D, Maung R, et al. A Pivotal Role for Interferon-α Receptor-1 in Neuronal Injury Induced by HIV-1. J Neuroinflammation 2020;17(1):226; doi: 10.1186/s12974-020-01894-2.

Sofroniew MV and Vinters HV. Astrocytes: Biology and Pathology. Acta Neuropathol 2010;119(1):7–35; doi: 10.1007/s00401-009-0619-8.

Sun B and Wang J. Guanylate Binding Protein (GBP) 4 Negatively Regulates Virus Induced Type I Interferon and Antiviral Response by Targeting Interferon Regulatory Factor (IRF) 7 (168.30). J Immunol 2012;188(1 Supplement):168.30-168.30.

Thaney VE, O'Neill AM, Hoefer MM, et al. IFNβ Protects Neurons from Damage in a Murine Model of HIV-1 Associated Brain Injury. Sci Rep 2017;7:46514; doi: 10.1038/srep46514.

Toggas SM, Masliah E, Rockenstein EM, et al. Central Nervous System Damage Produced by Expression of the HIV-1 Coat Protein Gp120 in Transgenic Mice. Nature 1994;367(6459):188–193; doi: 10.1038/367188a0.

Tripal P, Bauer M, Naschberger E, et al. Unique Features of Different Members of the Human Guanylate-Binding Protein Family. J Interferon Cytokine Res 2007;27(1):44–52; doi: 10.1089/jir.2007.0086.

Ulvestad E, Williams K, Bjerkvig R, et al. Human Microglial Cells Have Phenotypic and Functional Characteristics in Common with Both Macrophages and Dendritic Antigen-Presenting Cells. J Leukoc Biol 1994;56(6):732–740; doi: 10.1002/jlb.56.6.732.

UNAIDS. Global HIV & AIDS Statistics — 2021 Fact Sheet. 2022. Available from: https://www.unaids.org/en/resources/fact-sheet [Last accessed: 8/23/2022].

Vestal DJ and Jeyaratnam JA. The Guanylate-Binding Proteins: Emerging Insights into the Biochemical Properties and Functions of This Family of Large Interferon-Induced Guanosine Triphosphatase. J Interferon Cytokine Res 2011;31(1):89–97; doi: 10.1089/jir.2010.0102.

Villacampa, N., Almolda, B., González, B., & Castellano, B. (2013). Tomato Lectin Histochemistry for Microglial Visualization. In B. Joseph & J. L. Venero (Eds.), Microglia: Methods and Protocols (pp. 261–279). Humana Press

Woollard SM, Bhargavan B, Yu F, et al. Differential Effects of Tat Proteins Derived from HIV-1 Subtypes B and Recombinant CRF02_AG on Human Brain Microvascular Endothelial Cells: Implications for Blood-Brain Barrier Dysfunction. J Cereb Blood Flow Metab 2014;34(6):1047–1059; doi: 10.1038/jcbfm.2014.54.

Appendix A:

Genes used in PSEA:

Neurons: Neurofilament Medium (NCBI gene id 4741) and Glutamate Decarboxylase (Kodama et al., 2012);

Astrocytes: Glial Fibrillary acidic protein (Sofroniew and Vinters, 2010) and Aldehyde Dehydrogenase 1 Family Member L1 (Adam et al., 2012);

Microglia: CD68 (Ulvestad et al., 1994) and Allograft Inflammatory Factor 1 (Deininger et al., 2002);

Oligodendrocytes: Myelin Oligodendrocyte Glycoprotein (Scolding et al., 1989), Myelin Basic Protein (Barbarese, 1991) and Myelin Associated Glycoprotein (Quarles, 1983).

Illumina Probeset IDs used in PSEA:

Neuron: ILMN_1225279 (Nefm), ILMN_2632416 (Nefm), ILMN_2938820 (Nefm), ILMN_2621743 (Gad1)

Astrocyte: ILMN_1214715 (Gfap), ILMN_1215847 (Gfap), ILMN_3100276 (Aldh1l1)

Microglia: ILMN_2689785 (Cd68), ILMN_1212938 (Aif1), ILMN_2804487 (Aif1)

Oligodendrocyte: ILMN_1259536 (Mog), ILMN_2615034 (Mog), ILMN_1227299 (Mbp), ILMN_1237021 (Mag)

ACCT: a Fast and Accessible Automatic Cell Counting Tool Using Machine Learning for 2D Image Segmentation

Theodore J. Kataras[1,2], Tyler J. Jang[1,2], Jeffrey Koury[2], Hina Singh[2], Dominic Fok[2], and Marcus Kaul[1,2,*]

[1]University of California, Riverside, Graduate Program of Genetics, Genomics and Bioinformatics, 92507, USA

[2]University of California, Riverside, Dept. Biomedical Sciences, 92507, USA

ABSTRACT

Counting cells is a cornerstone of tracking disease progression in neuroscience. A common approach for this process is having trained researchers individually select and count cells within an image, which is not only difficult to standardize but also very time-consuming. While tools exist to automatically count cells in images, the accuracy and accessibility of such tools can be improved. Thus, we introduce a novel tool ACCT: Automatic Cell Counting with Trainable Weka Segmentation which allows for flexible automatic cell counting via object segmentation after user-driven training. ACCT is demonstrated with a comparative analysis of publicly available images of neurons and an in-house dataset of immunofluorescence-stained microglia cells. For comparison, both datasets were manually counted to demonstrate the applicability of ACCT as an accessible means to automatically quantify cells in a precise manner without the need for computing clusters or advanced data preparation.

Introduction

Quantifying cells in immunofluorescent images has long been a limiting step in both time and required effort for the analysis of microscopy data used in research. These selective image analysis techniques can provide valuable physiological information and manual counts by trained professionals have been held up as the "gold standard" for quantification (Jensen, 2013; Schneider et al., 2012; von Bartheld et al., 2016).

Here we used multiple separate observers' complete manual counts for comparison to an automatic cell counting methodology. Traditionally, an important aspect of maintaining consistency in cell quantification has been ensuring that a dataset is counted by a single observer who strives for accuracy and reproducibility while ideally being blinded to the experimental conditions. This massively limits the speed at which cell counting data can be processed, as increases in manpower do not always translate to increased speed. Manual counting can struggle with reproducibility and consistency across a dataset due to human error and fatigue. Such issues can be avoided by utilizing computational models which remain consistent over any number of images.

For that purpose we introduce here ACCT: Automatic Cell Counting with Trainable Weka Segmentation (TWS) is publicly hosted for download on GitHub at github.com/tkataras/Automatic-Cell-counting-with-TWS.git. TWS provides a machine learning basis for our accessible automatic cell counting methodology, with additional image processing potential provided by scripts in ImageJ, Python, and BeanShell (Arganda-Carreras et al., 2017; Schneider et al., 2012). The TWS program provides a graphical user interface (GUI) for training and applying a machine learning classifier that differentiates between cell and non-cell pixels,

which are then grouped into cell objects and counted. ACCT is built around this pixel segmentation to provide quantitative validation at the cellular level and assist in optimal classifier selection and application (Fig. 1).

Two datasets are used in this study to demonstrate performance in varied imaging contexts. The first dataset used is comprised of imaged microglia in mice with and without immune-and-inflammation-activating conditions brought on by the transgenic expression of the envelope protein gp120 of human immunodeficiency virus-1 (HIV-1) (Toggas et al., 1994). This model of NeuroHIV (HIVgp120tg mouse) provides an observable outcome from the manual counts, an increase in microglia in the presence of HIVgp120 (referred to hereafter as Activated) versus the absence of the viral protein (non-transgenic littermate control, referred to as Resting). ACCT was used to assess the difference in microglia cell numbers from images represented in Fig. 2. For an automatic counting methodology to be effective in an experimental context, it must be able to accommodate the variability in data presentation resultant from experimental conditions (Lynch, 2009). Microglia are known to undergo morphological changes during activation that alter their morphology and appearance when imaged through immunofluorescent staining (Gomez-Nicola & Perry, 2015; Karperien et al., 2013).

We focus on a dataset of images of cells immunofluorescence-labeled for ionized calcium-binding adaptor protein-1 (Iba-1) which is a cell type-specific marker and enables visualizing microglia. However, the methodology and accompanying scripts allow for automatic quantification of cells in a wide array of imaging contexts.

The second dataset used is a publicly available set of images of monosynaptic retrograde tracer stained neurons at 200x magnification (Fig. 3). This dataset, which we refer to as the Fluocell dataset, was used in the generation of novel additions to the U-net neuronal network for cell segmentation (Clissa, 2021; Morelli et al., 2021). We use this dataset to demonstrate ACCT under a different set of challenges on data produced outside our lab.

Accessibility

The existence of software tools for use in the life sciences does not inherently lead to an improvement in function (Goecks et al., 2010). The prerequisite technical knowledge to operate new software tools effectively can create barriers to novel methodologies based on their accessibility. The goal of ACCT is to reduce the barrier to entry for the execution of full semi-supervised imaging studies. ACCT provides the tools to leverage user expertise in handcrafting training data, while providing quantitative tools to efficiently assess training accuracy from a variety of approaches. By reducing the programming knowledge required from users with GUI elements, ACCT increases accessibility of automatic cell counting. Additionally, ACCT performs statistical analysis from the counted images, reducing the technical workload and additionally increasing the tool's accessibility.

Related Works

There are many ways to address an image segmentation problem. This complex problem centers on assigning an appropriate label for every pixel in an image. The TWS program we utilize is just one of several software tools, including machine learning implementations such as Ilastik (Berg et al., 2019) and neural nets like U-Net, ResUNet, and c-ResUnet (Diakogiannis et al., 2020; Morelli et al., 2021; Ronneberger et al., 2015).

We have chosen to work with TWS (Arganda-Carreras et al., 2017) over Ilastik (Berg et al., 2019) due to the increased breadth of default available features, as well as the integration with Fiji and ImageJ that streamlines automated image processing and analysis. This integration with ImageJ made TWS more accessible to build upon for this and future automatic imaging tools.

While programs like TWS and Ilastik provide excellent pixel segmentation with an accessible interface, there is an additional need to assess accuracy and performance at the cell level, rather than the pixel level. ACCT provides a framework for users to accomplish this with minimal file manipulation at the command line. Ilastik's segmentation does not test models against a validation stage following training of their machine learning model, which increases the risk of overfitting to the training dataset. Thus, we compare the performance of Ilastik to ACCT in our study.

Additionally, we compare the performance of ACCT against CellProfiler, which is a tool commonly used for image analysis which allows users to create modular pipelines (Stirling et al., 2021). This tool provides pixel level segmentation, although it does not provide automated cell counting with machine learning models without its companion tool CellProfiler Analyst (Dao et al., 2016). However, CellProfiler Analyst requires the users to manually modify text and database files in SQL, which requires user knowledge of code editors. For this reason, we do not compare against CellProfiler Analyst.

Finally, ResUNet is a convolutional neural net approach (CNN) to image segmentation and exists as a general tool for image labeling. It was demonstrated to make effective use of training data to make accurate cell segmentation on images with a large variance in the number of

cells, as well as the presence of non-cell artifacts. This is a development on U-Net, which has proven effective at bulk cell counting tasks in a variety of contexts. Further, c-ResUnet is an extension of ResUNet (Morelli et al., 2021). However, CNN models require high processing power to generate results in a reasonable amount of time, which may require accessing expensive computational centers. ACCT is designed to be efficiently functional on commercially available consumer laptops and computers.

Methods

Iba-1 Microglia Dataset

A dataset comprised of images of Iba-1 positive microglial cells was generated following procedures recently published by our group (Singh et al., 2020). In brief, the dataset was derived from brain sections of a mouse model for HIV-induced brain injury (HIVgp120tg), which expresses soluble gp120 envelope protein in astrocytes under the control of a modified GFAP promoter[5]. The mice were in a mixed C57BL/6.129/SJL genetic background, and two genotypes of 9 month old male mice were selected: wild type controls (Resting, n = 3) and transgenic littermates (HIVgp120tg, Activated, n = 3). No randomization was performed. HIVgp120tg mice show among other hallmarks of human HIV neuropathology an increase in microglia numbers which indicates activation of the cells compared to non-transgenic littermate controls (Singh et al. 2020). All experimental procedures and protocols involving animals were performed in compliance with National Institutes of Health (NIH) guidelines and approved by the Institutional Animal Care and Use Committees (IACUC) of the Sanford Burnham Prebys Medical Discovery Institute (SBP), The Scripps Research Institute (TSRI), and the University of California Riverside (UCR). The study follows ARRIVE guidelines.

Figure

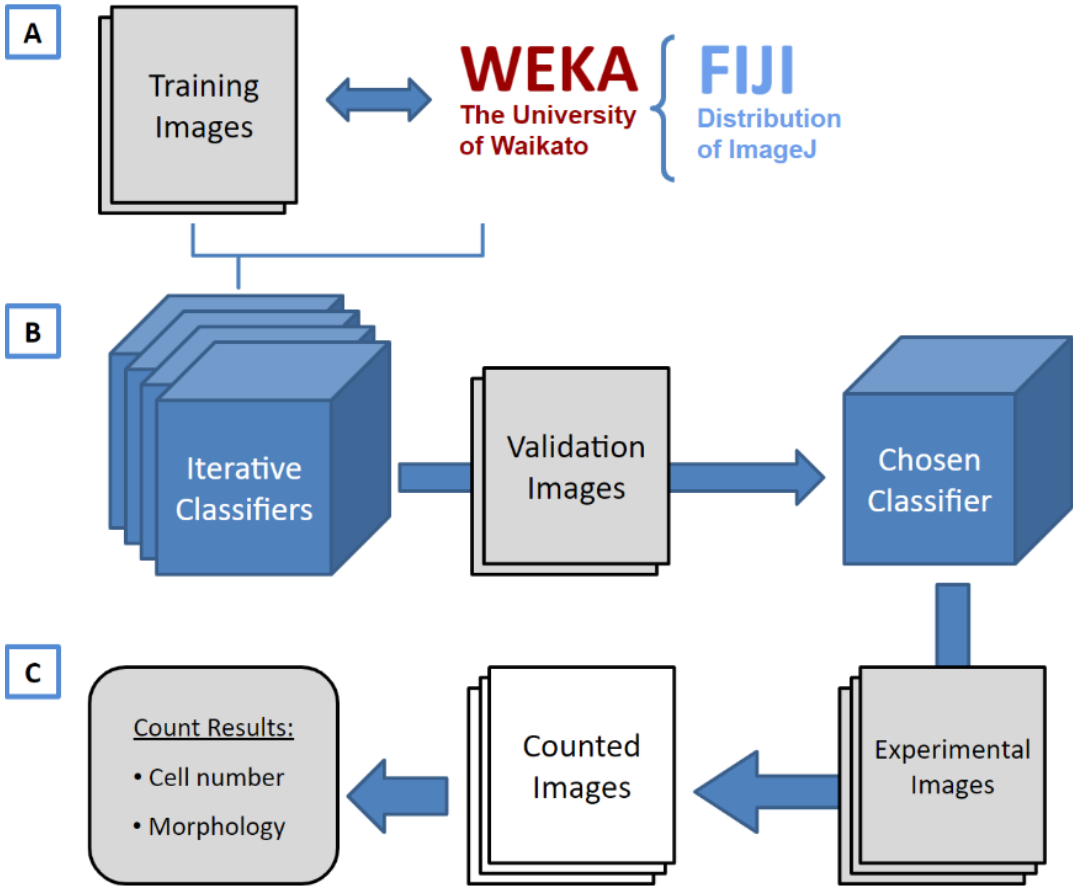


Fig. 1 A visual overview of ACCT components and process. (A) Weka and a set of training images are employed to create multiple classifiers through iterative training. (B) These classifiers are then evaluated in bulk against validation images and the best classifier is chosen by the user. (C) The chosen classifier is applied to the experimental dataset for cell quantification, producing a set of counted images, as well as information about the cell morphology.

The procedures for brain tissue harvest, immunofluorescence staining, and microscopy of microglia have been described in a recent publication by our group (Singh et al. 2020). In brief, mice were terminally anesthetized with isoflurane and transcardially perfused with 0.9% saline. The mouse brains were removed and fixed for 72 hours at 4°C in 4% paraformaldehyde. Brain sections were obtained using a vibratome (Leica VT1000S, Leica Biosystems, Buffalo Grove, IL) and cerebral cortex in 40 µm thick sagittal sections spaced 320 µm apart medial to lateral from brains of of each genotype. Staining was performed with rabbit anti-ionized calcium-binding

adaptor molecule 1 (Iba-1) IgG (1:125; Wako) with secondary antibody Fluorescein isothiocyanate (FITC). For quantification of Iba-1 stained microglia, cell bodies were counted in the cerebral cortex from three fields of view for three sections each per animal. Between 2 and 3 images were collected per field of view to capture as many cells as possible in sufficient focus for identification. Microscopy was performed with a Zeiss 200 M fluorescence deconvolution microscope with a computer-controlled 3D stage and FITC filter. All images were collected using Slidebook software (version 6, Intelligent Imaging Innovations, Inc., Denver, CO). Images were acquired at 10X magnification and pixel resolution 1280x1280 and cropped to 1280x733 pixel area to exclude irregular tissue edges. Representative examples are shown in Fig. 2.

Manual Counts

Manual counts were performed by three observers, who were allowed to adjust the image brightness to best facilitate counting accuracy. Images were collected as a Z-stack consisting of two to three planes of focus 0.5 μm apart per field in order to allow the observer to confirm the presence of Iba-1 positive cell bodies that were only partially in focus. The plane showing most cells in focus was used as the primary plane for counting. The observers used different visualization software during counting. Observer A used the Slidebook software (Intelligent Imaging Innovations, Denver, CO) paired with the microscope and Observers B and C used the Fiji distribution of ImageJ 2.1.0 for manual counting. Additionally, Observer A's count was performed prior to the start of this project, and count markers were placed on images in close proximity to cell bodies for rapid total summation. Observer B and C placed counts within cell bodies to allow for later cell-level accuracy assessment. Microglia counts were normalized to area in cases where a part of the image area was unsuitable for cell detection, due to tissue damage or thickness irregularity (n = 3 images out of 62 total in study).

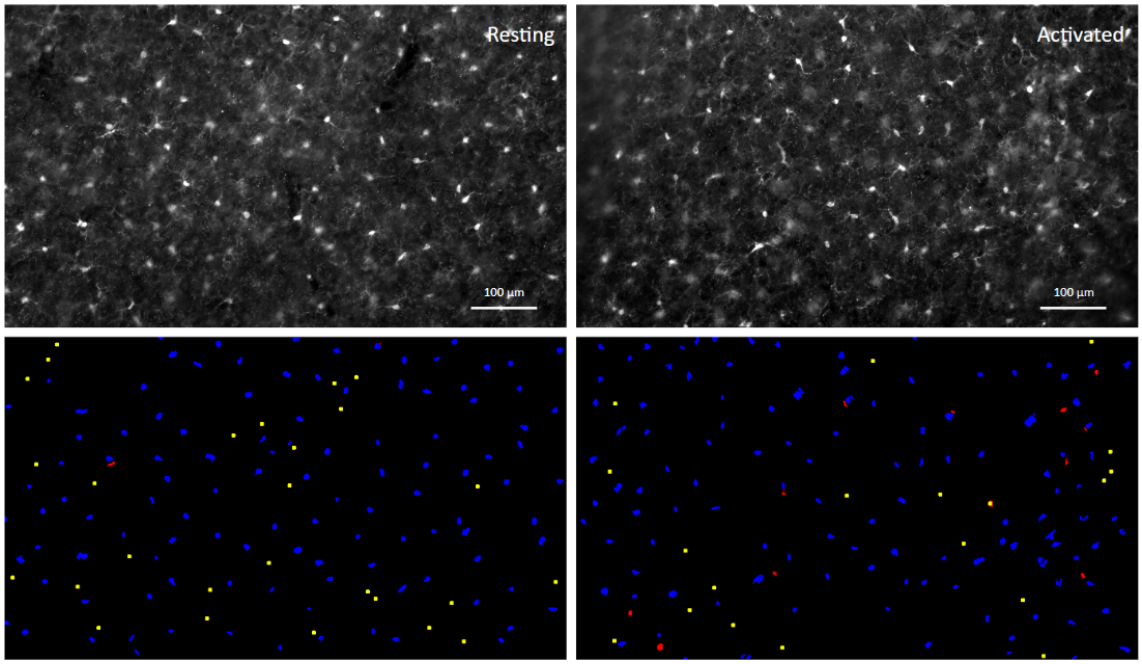


Figure 2. Images of immunofluorescence-labeled microglia before and after segmentation. An example of processed paired images of Iba-1 immunolabeled microglia in cerebral cortex (layer III; Upper panel) of wild-type, non-transgenic ('Resting') and of HIVgp120tg mice ('Activated'), and the accompanying final segmentation images generated via ACCT (lower panel). Resulting object segmentations are colored coded (blue = true positive, red = false positive, yellow = false negative). Segmented objects from images in the same field of view were projected with a size exclusion minimum of 50 pixels for counting. Immunofluorescence staining and acquisition of images are described in previous publications and the Methods Section. Scale bar: 100 μm.

Fluocell Public Dataset

To further examine and develop the effectiveness of ACCT on a variety of data, we also performed a cell counting study using a publicly available image set (Morelli et al. 2021; Clissa 2021). The 283 1600x1200 pixel images were taken at 200x magnification of 35 µm thick slices of mouse brain tissue with neurons stained via a monosynaptic retrograde tracer (Cholera Toxin b). This tracer highlighted only neurons connected to the toxin injection site.

This dataset contains images with both high and low cell density, as well as varying amounts of noise and artifacts (Supplementary Fig. S2). We also observed that many images contain overlapping or touching cells. The Fluocell dataset presents different challenges when compared to our Iba-1 positive microglia dataset where cells are more evenly distributed and the number of cells per image is more consistent. A representative example of Fluocell data is shown in Fig. 3.

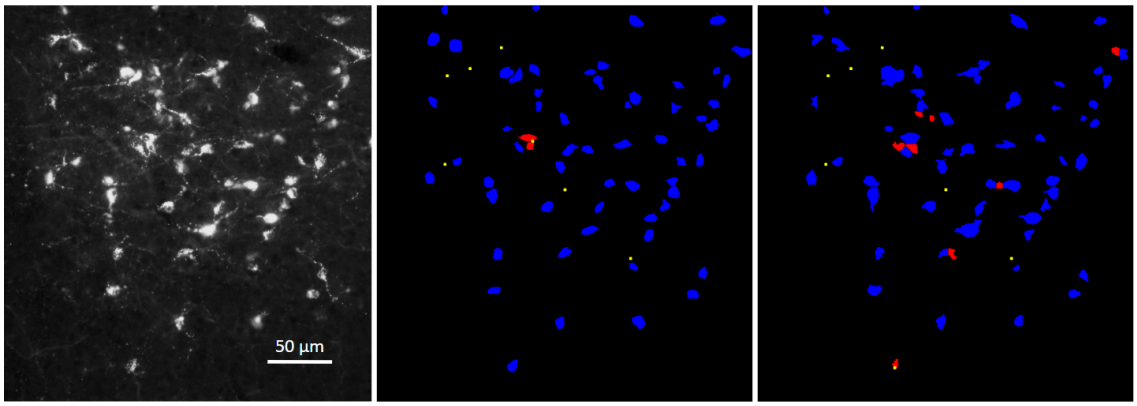
In the Fluocell analysis of this data, a subset of the images were manually counted by the authors, and the remaining images were counted via automatically thresholding[9]. Since we wish to compare ACCT to human placed cell counts as ground truth labels to assess performance of our tool versus human cell counting, we manually counted the entire 283 image dataset (one observer). This allows us to validate our tool against manual observer cell counts rather than another automatic process. In addition, the authors of the Fluocell dataset wrote their own automated cell counting program using a CNN approach named c-ResUnet which builds upon ResUNet[9]. Thus, we also compare the performance of ACCT versus c-ResUnet and Ilastik on the Fluocell dataset with our manual counts.

Automatic Counting Methodology

ACCT is open source at github.com/tkataras/Automatic-Cell-counting-with-TWS.git. Our machine learning classifier was built using the TWS plugin version 3.2.34 in ImageJ 2.1.0 included in the Fiji distribution[18]. In addition, the open source Python packages: scipy, pandas, numpy, matplotlib, imageio, and scikit-learn were used in ACCT (Virtanen et al. 2020; McKinney 2010; Harris et al. 2020; Hunter 2007; Kleine 2021; Pedregosa et al. 2011).

ACCT allows for the selection of several different types of machine learning approaches. Machine learning here refers to dynamic models trained on user specified input data to select cell pixels within an image. Users can also upload additional machine learning approaches compatible with Weka if desired. For this paper, we use an implementation of the Random Forest approach, called Fast Random Forest (Arganda-Carreras et al. 2017; Breiman 2001). This is the default machine learning approach in TWS and the following default features were used:

- Gaussian blur
- Sobel filter
- Hessian
- Difference of Gaussian
- Membrane Projections
- Membrane thickness = 1
- Membrane patch size = 19
- Minimum sigma = 1.0
- Maximum sigma = 16.0



(a) Original Image (b) c-ResUnet Segmentation (c) ACCT Segmentation

Figure 3. Image of fluorescence-labeled neurons before and after segmentation. A cropped image taken from the Fluocell dataset paired with cell segmentation (a). This depicts segmented cell objects from image MAR38S1C3R1_DMR_20_o reported in the publicly available Fluocell dataset segmented by c-ResUnet[10] (b) and ACCT with classifierBayes3 (c). Resulting object segmentations are colored coded (blue = true positive, red = false positive, yellow = false negative). ACCT was set to filter out objects smaller than 250 pixels and greater than 5,000 pixels to remove noise. We applied the watershed algorithm to this dataset. ACCT correctly identified 84.6% of the hand counts in this image and c-ResUnet 86.2%, while ACCT was correct with 86.9% of all predictions and c-ResUnet with 93.3%. The scale bar represents a length of 50 µm.

We additionally use a Bayesian Network model, which is also implemented in Weka[4]. This approach, called BayesNet, follows a Bayesian statistical model to determine the probability that observed features are conditionally dependent, or caused, by the object of interest[26]. For this study, we use the following parameters for Bayesian pixel classification, in addition to the above listed features:

- Variance
- Mean
- Minimum
- Maximum
- Median
- Anisotropic Diffusion
- Bilateral
- Lipschitz
- Kuwahara
- Gabor
- Entropy
- Neighbors

During cell detection, small and large cellular processes or artifacts can be classified as cell bodies given similar enough appearance to cells. We address this noise by implementing a minimum and maximum cell object size parameter when counting cells. Thus, objects outside the specified size range are excluded from the automatic count. This range is empirically determined by observed cell bodies during model training and validation.

An additional challenge for cell detection is when two or more cells abut or overlap. This causes multiple cells to be identified as one large cell, so ACCT must separate these objects to increase accuracy. Thus, we optionally enable a watershed algorithm post pixel segmentation (Vincent and Soille 1991). This algorithm is used to separate objects by contour, which allows for separated objects to be counted independently. We use the default implementation of the watershed algorithm provided in ImageJ.

Cell Body Detection

The machine learning models in TWS generate a probability map for each image which is a representation of each pixel in the image as the probability that it is part of an object of interest. This probability is compared to a confidence threshold, which is the minimum probability a pixel must be to be considered part of an object. The user can set different threshold values, which affects how conservative or liberal the program will be in identifying objects. By default, ACCT starts at a threshold of 0.5 which can be modified by users through the user interface. Conventionally, stricter thresholds lead to fewer false positives but also fewer true positives. The inverse also holds with a more relaxed threshold identifying more true positives, but also more false positives. The performance of ACCT at various thresholds is represented visually on a Receiver Operator Characteristic (ROC) curve. However, some models usable with TWS in ACCT only give a binary zero or one for their confidence values which prevents generation of meaningful ROC curves.

Iterative Training and Validation

Training on the Iba-1 microglia dataset was drawn from 10 randomly selected images not used in the counting analysis. These images were collected using the above described methods from mice distributed equally between experimental genotypes (Fig. 2). Incremental adjustments to training data and resulting changing pixel classification was observed in real time and the classifiers were saved sequentially.

To avoid overtraining, classifiers are updated a few pixels at a time with new training data and the updated pixel segmentations on training data are observed immediately in TWS. Subsequent training data is selected to address areas of incorrect segmentation. We continue this

over successive iterations of classifiers, saving a version of the classifier after each addition of training data. This iterative classifier creation scheme continues until the classifier does not appear to be improving on the data. ACCT then performs accuracy assessment on validation data and ground truth markers, accounting for experimental conditions, to help the user select the classifier iteration with the greatest accuracy and consistency across the validation dataset.

Multiple sequential classifiers were applied to the validation dataset (Fig. 1, Supplementary Fig. S1). Ultimately, 25 classifiers were trained on the Iba-1 microglia data. The Iba-1 microglia validation dataset was comprised of 10 images (Resting n = 5, Activated n = 5) from the main dataset which was then excluded from all further analyses. Cell body location specific count markers were placed in these images by Observer B, and performance was calculated via precision, recall, F1 score, accuracy, as well as a Student's T-Test of differential accuracy between the Resting and Activated images. ACCT is also able to perform ANOVA calculations for further analyses including more than two experimental groups.

In the Fluocell data, 10 training images were selected from the dataset to represent the variety of segmentation challenges within the dataset: highly variable intensity, highly variable cell density, overlapping cell images, and images with non-cell artifacts. The validation dataset was made with a different set of 10 images selected to represent a similar distribution of challenges.

Classifier Selection

True positive scores for each image are determined by the localization of manual count markers which are checked against the pixel location values of each object for the automatic counting process. False positives for each image are represented as the total number of automatically generated cell objects that do not contain a single manual count. False negatives for each image are determined by the total number of manual count markers that are not contained inside of an automatically generated object by the program, plus the number of manual count markers inside of a single cell object in excess of one, indicating insufficient object separation. As we assess accuracy based on cell location and do not differentiate between background pixels, ACCT does not include determination of true negative cell locations (Morelli et al. 2021). Thus, we calculate accuracy as:

$$\frac{TP}{TP+FP+FN}$$

Where:

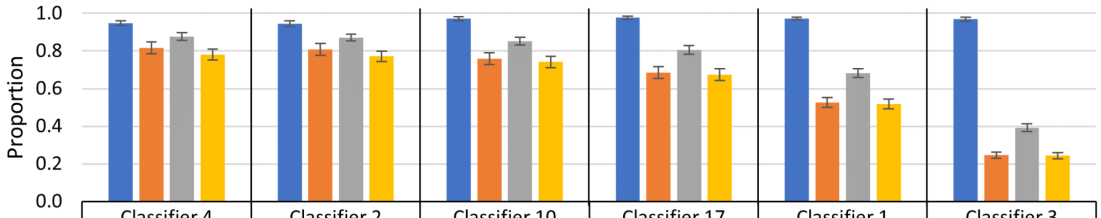
TP = true positive = a cell object with a hand placed marker inside

FP = False positive = a cell object lacking a hand placed marker

FN = false negative = a hand placed marker outside of any cell object

We assess the performance of our classifiers using measures of precision, recall, F1 score, and accuracy. Precision is the proportion of automatic counts that are correct based on manually placed markers, and recall is the proportion of the total manually placed cell markers that were successfully identified by the automatic count. The F1 score is the harmonic average of precision and recall. The accuracy is assessed specifically as the number of true positive cell counts as a proportion of all manual and automatic counts, including false negatives. Multiple classifiers can be evaluated through automatically calculated statistical analysis. Statistical measures, such as

mean absolute error (MAE), are additionally calculated through ACCT to evaluate the performance of different classifiers. We used these statistics to assess the performance of ACCT against other automatic cell counting tools based on these metrics.



	Classifier 4	Classifier 2	Classifier 10	Classifier 17	Classifier 1	Classifier 3
Precision	0.949	0.945	0.971	0.977	0.973	0.969
Recall	0.815	0.808	0.759	0.685	0.526	0.246
F1	0.877	0.871	0.852	0.805	0.683	0.392
Accuracy	0.781	0.772	0.742	0.674	0.519	0.244

Precision Recall F1 Accuracy

Fig. 4 Summary of individual classifier performance on the Iba-1 microglia dataset during the validation stage. A chart of the most and least accurate three incrementally trained classifiers, ranked by F1 score of 25 trained classifiers (n = 10 images). Error bars represent standard error of the mean on calculated performance statistics from each image, where the statistic itself is calculated from total cells in the dataset. The parameters used in ACCT: 0.5 threshold, 50 minimum pixel size, and 1,000 maximum pixel size.

This statistical information is shown in Fig. 4 and Fig. 8. Fig. 4 is a subset of the full data which can be found in Supplementary Table T1. Fig. 8 shows selected accuracy statistics at different confidence thresholds. Classifier selection via the best F1 score or different weighting of precision and recall, are all valid metrics for selecting a classifier. However, for this study we have selected the classifier based on the highest F1 score.

Experimental Dataset Analysis

As the next step, the selected classifier is applied to the experimental dataset of images. This experimental dataset excludes images used in training and validation. The automated counting methodology is repeated in this analysis and reports the total number of cells counted in

addition to other statistical information per image. Morphological information about each identified cell is reported by ACCT to users. An example of this can be found in Supplementary Table T2, which lists some of the reported morphological information generated from analysis.

Selected Classifier Performance Audit

The audit requires further manual counting and is identical to how we assess the performance of classifiers during the validation stage. This step is intended to determine how similarly the classifier performed on the experimental dataset compared to the validation set in situations where all images have not been manually counted. The audit can be performed using a subset of the experimental dataset, or even the whole dataset, if the user chooses to complete an entire manual count to assess model accuracy. These images are known as the audit set. We randomly selected 5 images each of Activated and Resting microglia experimental images for the Iba-1 audit set. An audit of the Fluocell data was performed on a sample of 10 images from the Fluocell experimental dataset.

Results

Iba-1 Microglia Dataset

Microglia Density

All statistical tests on Iba-1 microglia images include all images except those used for training and validation (Resting n = 22, Activated n = 20). We report classifier 10's automatic count instead of classifier 4 because classifier 10 provided the maximum performance on the experimental dataset and audit set that we observed. The significant increase in microglia density in images of gp120 positive (Activated) mice was consistent across the dataset via two way ANOVA ($p = 3.39E^{-16}$; Resting n = 22, Activated n = 20) (Fig. 5). Accounting for genotype variance within the dataset, a trend for decreases in mean microglia density were found between

the automated count and Observer A, but no significant differences were found between the ACCT count and Observers B and C (Observer A $p = 0.060$; Observer B $p = 0.514$; Observer C $p= 0.440$). Additionally, the per image microglia density demonstrated significant correlations among the automatic count and all observers with stronger correlations in Resting microglia images (Table 1). Visual representation of the data is found in Supplementary Fig. S3 as scatter plots with regression lines. The following represents the total cell counts:

- Validation: Observer A/B 1263/1380 cells over 10 images.
- Experimental Dataset: Observer A/B/C 5158/5035/5056 cells over 42 images.
- Audit Set: Observer A/B/C 1239/1207/1262 cells over 10 images.

Precision and Recall

The overall precision and recall achieved by the TWS methodology were similar in the validation dataset, experimental dataset, and audit dataset with overall accuracy and F1 increased in the experimental dataset compared to validation as shown for classifier 10 in Fig. 6. However, within the experimental dataset, the TWS classifier was more conservative in the Resting images compared to Observer B's manual counts, with the automatic count having higher precision in images of Resting than Activated samples (Precision $p = 0.007477$) (Fig. 6). When compared to Ilastik and CellProfiler, ACCT had a similar, and slightly stronger performance than both tools in each set of Iba-1 images, with Ilastik slightly outperforming CellProfiler. We additionally compared these tools against basic functionality that users can manually select in Fiji, to illustrate how ACCT builds upon existing Fiji functionality. We used the subtract background with rolling ball, adjust threshold, and watershed tools in Fiji for this analysis, with background subtraction at 25 pixel area and pixel intensity threshold of 90. Without applying minimum and maximum

object size this analysis resulted in near 0 precision, thus we used a 50 minimum and 1000 maximum pixel size as in the other tools. Classifier 10 outperforms the basic Fiji tools in most metrics except for precision. The basic Fiji application also narrowly outperforms Ilastik and CellProfiler in most metrics in the Iba-1 images.

	Resting	Resting	Activated	Activated
Correlation	p val.	Adj. R^2	p val.	Adj. R^2
Classifier 10 vs Observer A	$6.49E^{-}8$	0.764	$6.01E^{-4}$	0.4606
Classifier 10 vs Observer B	$2.42E^{-7}$	0.7313	$4.19E^{-5}$	0.5944
Classifier 10 vs Observer C	$3.30E^{-5}$	0.5653	$1.11E^{-4}$	0.5498
Observer B vs Observer A	$7.38E^{-6}$	0.6243	$4.15E^{-4}$	0.4814
Observer C vs Observer A	$2.05E^{-4}$	0.4816	$9.04E^{-7}$	0.7327
Observer B vs Observer C	$7.14E^{-9}$	0.8103	$6.28E^{-4}$	0.4581

Table 1. Correlation analysis of microglia density. The automatic count correlated with manual counts in both experimental conditions as well as correlations between manual counts, showing similar variability via correlation between the observer counts (Resting n = 22, Activated n = 20).

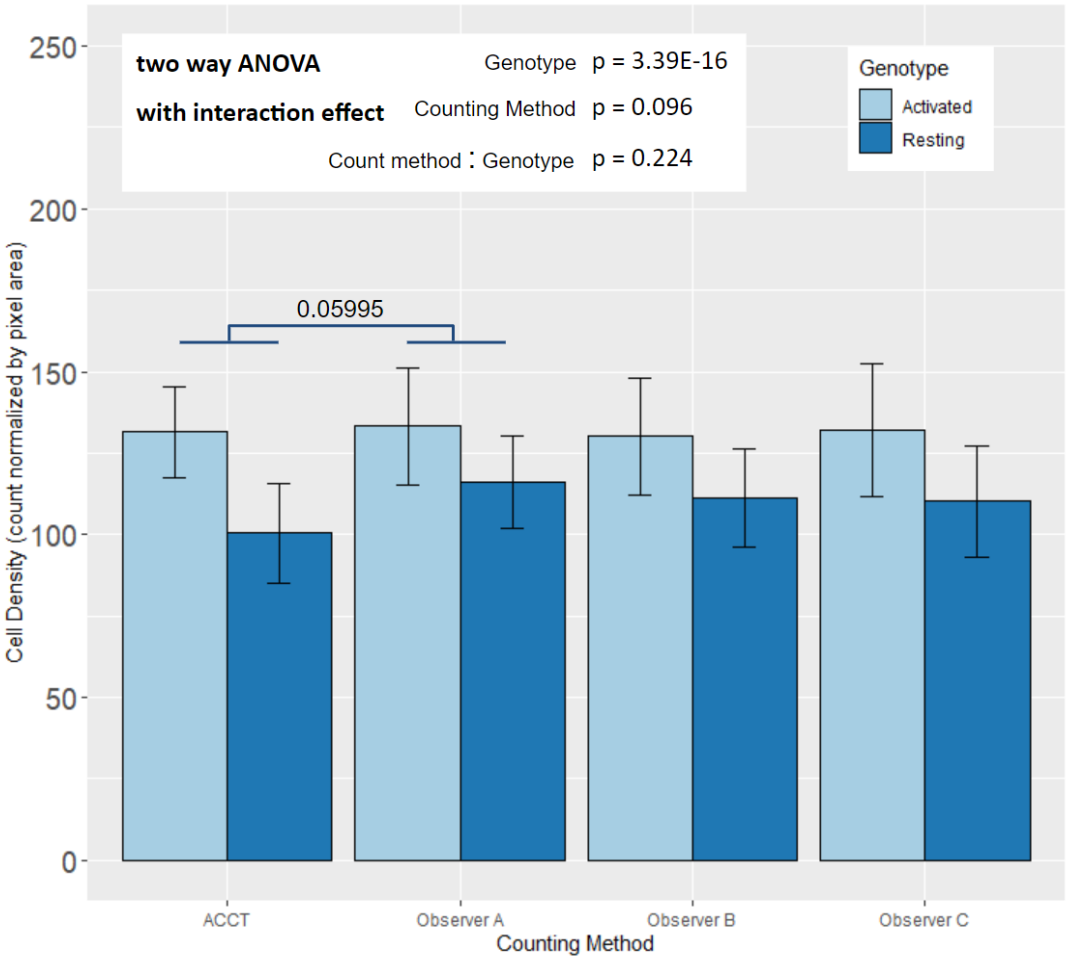
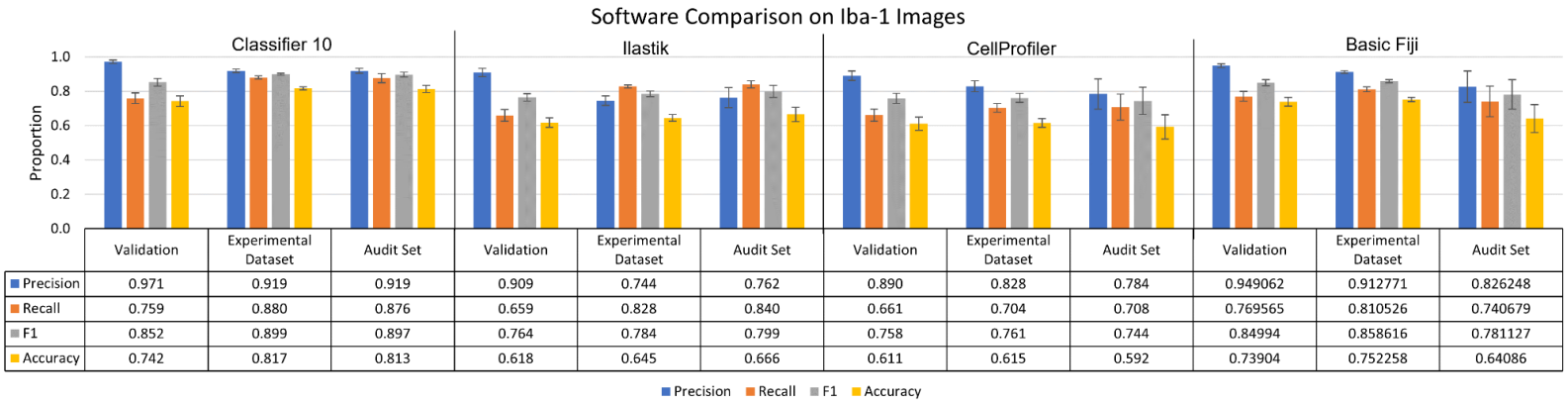


Fig. 5 Mean microglia density by experimental genotype in manual and automated counts. All counting methods found an increase in microglia density in Activated microglia images by two way ANOVA with interaction effect and Tukey HSD post-hoc analysis. The difference between counting method and the interaction effect did not display statistical significance. (Genotype: p = $3.39E^{-16}$; Counting Method: p = 0.096; Genotype:Counting Method: p = 0.224; Resting n = 22, Activated n = 20). There were not significant differences between the automatic count and Observer B and C. However, the ACCT count density trended lower overall than Observer A's (p = 0.0599; Resting n = 22, Activated n = 20). This suggests that classifier 10 may have excluded some faintly stained or not well-in-focus cells in the Resting group images that Observer A did count. Error bars represent standard deviation.

Fluocell Dataset

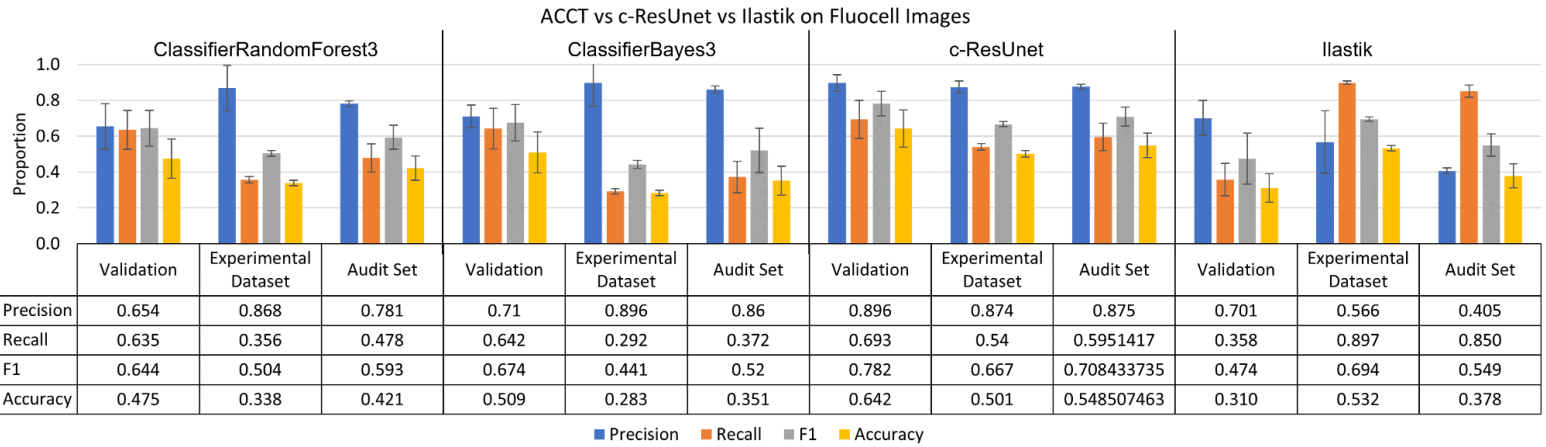
In contrast to the Iba-1 microglia dataset, the Fluocell dataset does not compare two different experimental conditions. All Fluocell statistical tests include all Fluocell images except those used in training and validation (n = 263). In Fig. 7, we compared the performance of the Fast Random Forest and BayesNet models implemented within ACCT versus c-ResUnet and Ilastik (Morelli et al. 2021; Berg et al. 2019). Fig. 7 shows ClassifierRandomForest3 outperforming BayesNet on most statistical metrics. Additionally, the c-ResUnet model outperformed on most metrics compared to these two classifiers. In contrast to the other tools, Ilastik has much greater recall than precision in the experimental and audit datasets, with ACCT and c-ResUnet outperforming on precision. However, Ilastik has the greatest F1 score in the experimental dataset. In the context of this dataset, the following represents the total cell count:

- Validation: 137 cells over 10 images.
- Experimental Dataset: 3307 cells over 263 images.
- Audit Set: 247 cells over 10 image



	Classifier 10			Ilastik			CellProfiler			Basic Fiji		
	Validation	Experimental Dataset	Audit Set	Validation	Experimental Dataset	Audit Set	Validation	Experimental Dataset	Audit Set	Validation	Experimental Dataset	Audit Set
Precision	0.971	0.919	0.919	0.909	0.744	0.762	0.890	0.828	0.784	0.949062	0.912771	0.826248
Recall	0.759	0.880	0.876	0.659	0.828	0.840	0.661	0.704	0.708	0.769565	0.810526	0.740679
F1	0.852	0.899	0.897	0.764	0.784	0.799	0.758	0.761	0.744	0.84994	0.858616	0.781127
Accuracy	0.742	0.817	0.813	0.618	0.645	0.666	0.611	0.615	0.592	0.73904	0.752258	0.64086

Fig. 6 ACCT vs Ilastik vs CellProfiler vs Manual Fiji on images of Iba-1 positive microglia. The performance of ACCT classifier 10 versus Ilastik, CellProfiler, and basic use of Fiji tools. The audit set is a selection of 10 images, which is equal to the size of the validation



	ClassifierRandomForest3			ClassifierBayes3			c-ResUnet			Ilastik		
	Validation	Experimental Dataset	Audit Set	Validation	Experimental Dataset	Audit Set	Validation	Experimental Dataset	Audit Set	Validation	Experimental Dataset	Audit Set
Precision	0.654	0.868	0.781	0.71	0.896	0.86	0.896	0.874	0.875	0.701	0.566	0.405
Recall	0.635	0.356	0.478	0.642	0.292	0.372	0.693	0.54	0.5951417	0.358	0.897	0.850
F1	0.644	0.504	0.593	0.674	0.441	0.52	0.782	0.667	0.708433735	0.474	0.694	0.549
Accuracy	0.475	0.338	0.421	0.509	0.283	0.351	0.642	0.501	0.548507463	0.310	0.532	0.378

Fig. 7 ACCT vs c-ResUnet vs Ilastik on the Fluocell dataset. ClassifierRandomForest3 and ClassifierBayes3 are the third trained iterations of a Fast Random Forest and BayesNet model in ACCT, respectively. The three tools' automatic counts of the Fluocell images are compared to our manual count of the Fluocell dataset. Error bars represent standard error of the mean on calculated performance statistics from each image, where the statistic itself is calculated from total cells in the dataset. The audit set is a selection of 10 images, which is equal to the size of the validation image set, chosen from the experimental dataset. The parameters used are: a 0.5 threshold, 250 minimum pixel size, 5,000 maximum pixel size, and with the watershed algorithm applied.

Receiver Operator Characteristic

ACCT automatically generates ROC curves for each trained classifier. This visualizes the tradeoffs between precision and recall as well as the true positive rate and the false positive rate. Fig. 8 demonstrates a ROC curve of ACCT classifier 10 applied to the Iba-1 microglia dataset. The threshold represents the required probability from the classifier to determine if a pixel will be designated as a cell pixel. The data represented in these graphs were generated using the scikit-learn Python library which performed statistical analysis (Pedregosa et al. 2011).

Fig. 8 demonstrates the tradeoff in which the false positive rate decreases at a faster rate than the true positive rate when a higher threshold is applied. For example, increasing the threshold for pixel segmentation in the Iba-1 dataset reduced the false positive rate compared to the default of 0.5. In this study, the 0.5 threshold was used for reported calculations, as overall accuracy did not increase due to a decrease in true positive cell identifications.

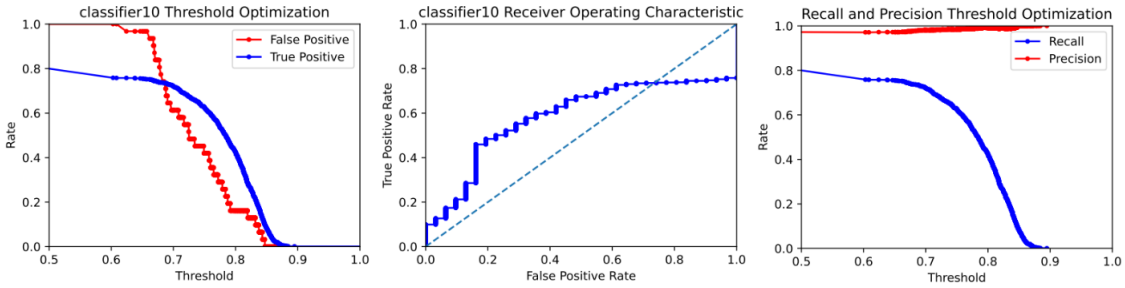


Fig. 8 An ROC curve generated by ACCT following the validation stage on Iba-1 microglia images. It depicts the false positive, true positive, recall, and precision rates of classifier 10 at different confidence thresholds on the Iba-1 stained microglia images. Objects were filtered to a minimum size of 50 pixels and a maximum size of 1,000 pixels. The watershed algorithm was not applied to the Iba-1 dataset, due to consistent cell separation in the sample tissues.

Discussion

ACCT is a step towards more accessible computation tools for cell counting and image segmentation. The main current advantage of this strategy is the shorter time required to train and apply the automatic counting strategy compared to manually counting each image. Our study demonstrates the general applicability of this tool to quickly explore large amounts of data.

The training process is critical for the success of this automatic cell counting methodology and relies on a researcher's specific knowledge of their imaged cell type. Each image set comes with its own unique set of challenges due to variability in cell and media characteristics, so providing accurate training data requires a firm and consistent understanding of the images in question. ACCT is able to help users adjust for these features in their images. Users can select specific features to be analyzed in their selected machine learning models to better represent their image data. Since every user has different data, this added flexibility improves the ability for ACCT to analyze users-specific images.

The results demonstrate that ACCT performs strongest on precision, indicating that most of the 'called' cells were real cells. Recall tends to be substantially lower than precision, leading to decreased F1 scores and accuracy in all ACCT classifiers tested on these datasets. This indicates that these models tend to be more conservative than expert manual counts. However, the results indicate that when models classify an object as a cell, they tend to be correct based on the high precision.

Additionally, microglia density analysis in Fig. 5 and in Table 1 demonstrate that ACCT counts cells similarly to expert observers. Counts of all observers identified a similar mean

difference in microglia density between experimental genotypes. ACCT correlates strongly with human cell counting results and can replicate the difference between experimental conditions similar to manual counting. Thus, it is a useful tool for image analysis between multiple experimental conditions.

We acknowledge that in the future more accurate automatic cell counting tools are likely to evolve from ACCT or other software packages. However, currently ACCT shows strong performance while being a more accessible tool to researchers than other approaches which require large computer networks or computing clusters. In some cases ACCT outperforms other cell counting tools, but not in every dataset. However, users may find the accessibility in the ImageJ environment and speed of ACCT worth trading for the slight loss in performance in some image data sets.

In terms of computational power, all work was performed on commercially available consumer laptops such as a Dell Inspiron 15-7559 (released Feb. 2017). Other automatic cell counting tools are often designed to make use of large computational resources. For example, Morelli et al. used 4 V100 GPUs to process their CNN approach (Morelli et al. 2021). CNN based tools recommend using a cluster, or network of multiple computers, which allows access to greater computational power. However, computer clusters are not available for all researchers and they additionally may require knowledge of command lines for effective utilization. ACCT is not limited by substantial computer specifications, additionally it is more accessible to less command line-oriented researchers.

Reproducible results are a major concern in scientific research, and ensuring reproducibility via manual cell counting can be costly and time consuming. Since ACCT stores

classifiers as single model files, they are easy to share and download. Thus, researchers can share reproducible results and statistical analysis of a cell counting study by sharing the model file and set of analyzed images. Since analysis generated by ACCT is stored as files editable in Excel, it is easy for users to share and communicate their results.

Building ACCT around the graphical interface implemented by TWS broadens usability by providing infrastructure for quantitative classifier validation and application in a full experimental context onto the flexible and intuitive training apparatus[4]. Since ACCT makes use of existing tools for cell imaging analysis such as TWS, researchers familiar with the program should find it easier to learn how to use ACCT as well.

ACCT includes accessible documentation, with an instruction manual that explains the program's function and usage found on its GitHub page. Documentation is important for users to understand and learn how to use software tools. Many other software tools document their components' functions inside of the tool itself, requiring users to navigate code to understand and use the tool. This is avoided by having detailed instructions written on the website for ACCT which explain the use of the tool without ever requiring the users to manually access the code itself.

ACCT could also be applied more generally to image segmentation problems. While the focus of our study is on cell counting in the context of neuroscience, so long as an image has object characteristics that can be separated from its background, ACCT is able to quantify the objects. However, more complex object shapes and less distinctive backgrounds may require selecting more complex models than demonstrated in this study. We provide a simple example of this in Supplementary Fig. S4 using the Fast Random Forest model, which is an image segmented for buildings against a field (Yan et al. 2021). While not as distinct as cells, it demonstrates that ACCT is applicable beyond the biological context. Overall, ACCT should greatly increase the accessibility of automatic analysis involving cell counting for a wide audience in neuroscience research and beyond.

Data Availability

The Iba-1 dataset, its analysis, and the Fluocell dataset analysis during the current study is available in the ACCT-Data Repository, https://github.com/tkataras/ACCT-Data-Repository.git.

The Fluocell dataset analyzed is publicly available as published in http://amsacta.unibo.it/6706/.

Code Availability

ACCT can be accessed and downloaded from GitHub at https://github.com/tkataras/Automatic-Cell-counting-with-TWS.git.

Acknowledgements

Funding from the National Institute of Health NIH, R01 MH104131, MH105330, MH087332, DA052209 to MK.

Author Contributions

T.K. and M.K. conceived of this project. T.J. and T.K. programmed the tool, performed the image analysis study, and wrote the manuscript. J.K. participated in image acquisition. User testing of the tool was done by T.K., T.J., and D.F. who provided feedback during ACCT development. Iba-1 microglia images by H.S. and manually counted by T.K., H.S., and D.F.. M.K. supervised the project and edited the manuscript. All authors reviewed the manuscript and approved of the manuscript before submission.

Additional information

Correspondence should be addressed to M.K. and requests for materials/tools should be addressed to T.K., T.J. or M.K.

Competing interests

The authors declare no competing interests.

References

Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K. W., Schindelin, J., Cardona, A., & Sebastian Seung, H. (2017). Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* , *33*(15), 2424–2426.

Berg, S., Kutra, D., Kroeger, T., Straehle, C. N., Kausler, B. X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., Eren, K., Cervantes, J. I., Xu, B., Beuttenmueller, F., Wolny, A., Zhang, C., Koethe, U., Hamprecht, F. A., & Kreshuk, A. (2019). ilastik: interactive machine learning for (bio)image analysis. *Nature Methods*, *16*(12), 1226–1232.

Clissa, L. (2021). *Fluorescent Neuronal Cells* [Data set]. University of Bologna. https://doi.org/10.6092/unibo/amsacta/6706

Dao, D., Fraser, A. N., Hung, J., Ljosa, V., Singh, S., & Carpenter, A. E. (2016). CellProfiler Analyst: interactive data exploration, analysis and classification of large biological image sets. *Bioinformatics* , *32*(20), 3210–3212.

Diakogiannis, F. I., Waldner, F., Caccetta, P., & Wu, C. (2020). ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing* , *162*, 94–114.

Goecks, J., Nekrutenko, A., Taylor, J., & Galaxy Team. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, *11*(8), R86.

Gomez-Nicola, D., & Perry, V. H. (2015). Microglial dynamics and role in the healthy and diseased brain: a paradigm of functional plasticity. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *21*(2), 169–184.

Jensen, E. C. (2013). Quantitative analysis of histological staining and fluorescence using ImageJ. *Anatomical Record* , *296*(3), 378–381.

Karperien, A., Ahammer, H., & Jelinek, H. F. (2013). Quantitating the subtleties of microglial morphology with fractal analysis. *Frontiers in Cellular Neuroscience*, *7*, 3.

Lynch, M. A. (2009). The multifaceted profile of activated microglia. *Molecular Neurobiology*, *40*(2), 139–156.

Morelli, R., Clissa, L., Amici, R., Cerri, M., Hitrec, T., Luppi, M., Rinaldi, L., Squarcio, F., & Zoccoli, A. (2021). Automating cell counting in fluorescent microscopy through deep learning with c-ResUnet. *Scientific Reports*, *11*(1), 22920.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675.

Singh, H., Ojeda-Juárez, D., Maung, R., Shah, R., Roberts, A. J., & Kaul, M. (2020). A pivotal role for Interferon-α receptor-1 in neuronal injury induced by HIV-1. *Journal of Neuroinflammation*, *17*(1), 226.

Stirling, D. R., Swain-Bowden, M. J., Lucas, A. M., Carpenter, A. E., Cimini, B. A., & Goodman, A. (2021). CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics*, *22*(1), 433.

Toggas, S. M., Masliah, E., Rockenstein, E. M., Rall, G. F., Abraham, C. R., & Mucke, L. (1994). Central nervous system damage produced by expression of the HIV-1 coat protein gp120 in transgenic mice. *Nature*, *367*(6459), 188–193.

von Bartheld, C. S., Bahney, J., & Herculano-Houzel, S. (2016). The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *The Journal of Comparative Neurology*, *524*(18), 3865–3895

Supplementary Figures

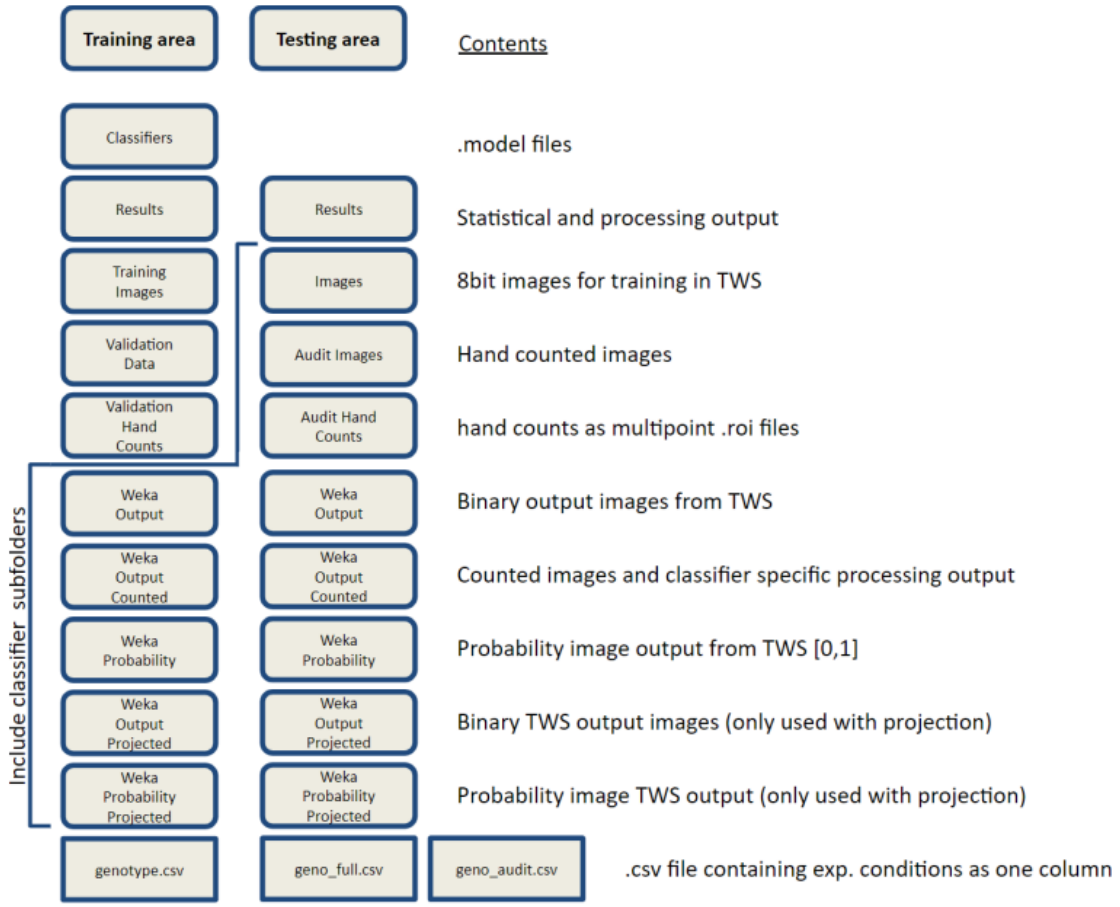


Figure S1. Map of the files and folders included in the ACCT pipeline. Folders are found under two primary folders: training area and testing area. The only folders the user will need to interface with directly are: Classifiers, the Results folders, Validation Data, Validation hand counts, Images, Audit images and Audit Hand counts

.

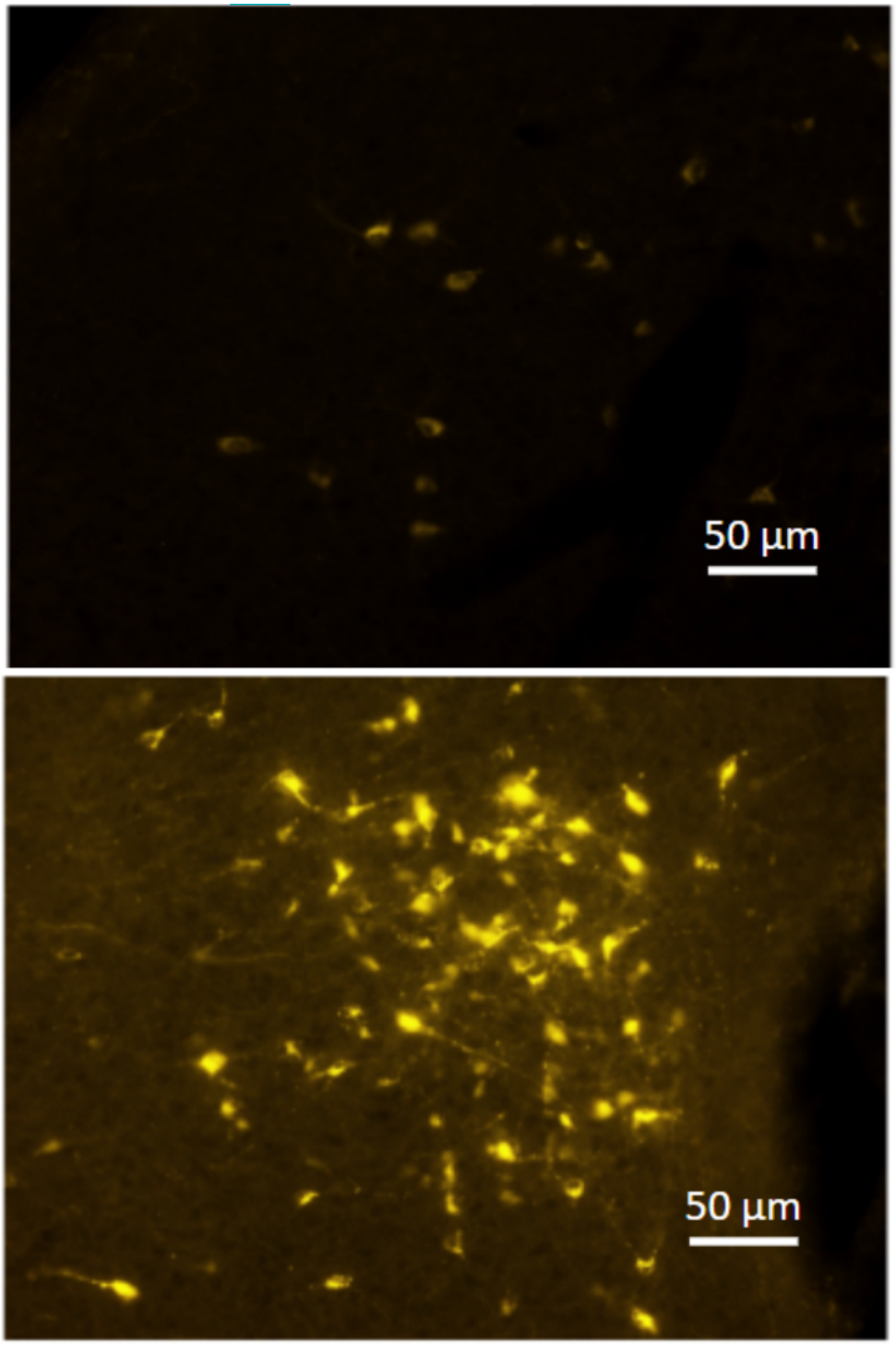


Figure S2. A comparison of the differences in images contained in the Fluocell dataset. These images demonstrate the challenging variance in cell presentation and density.

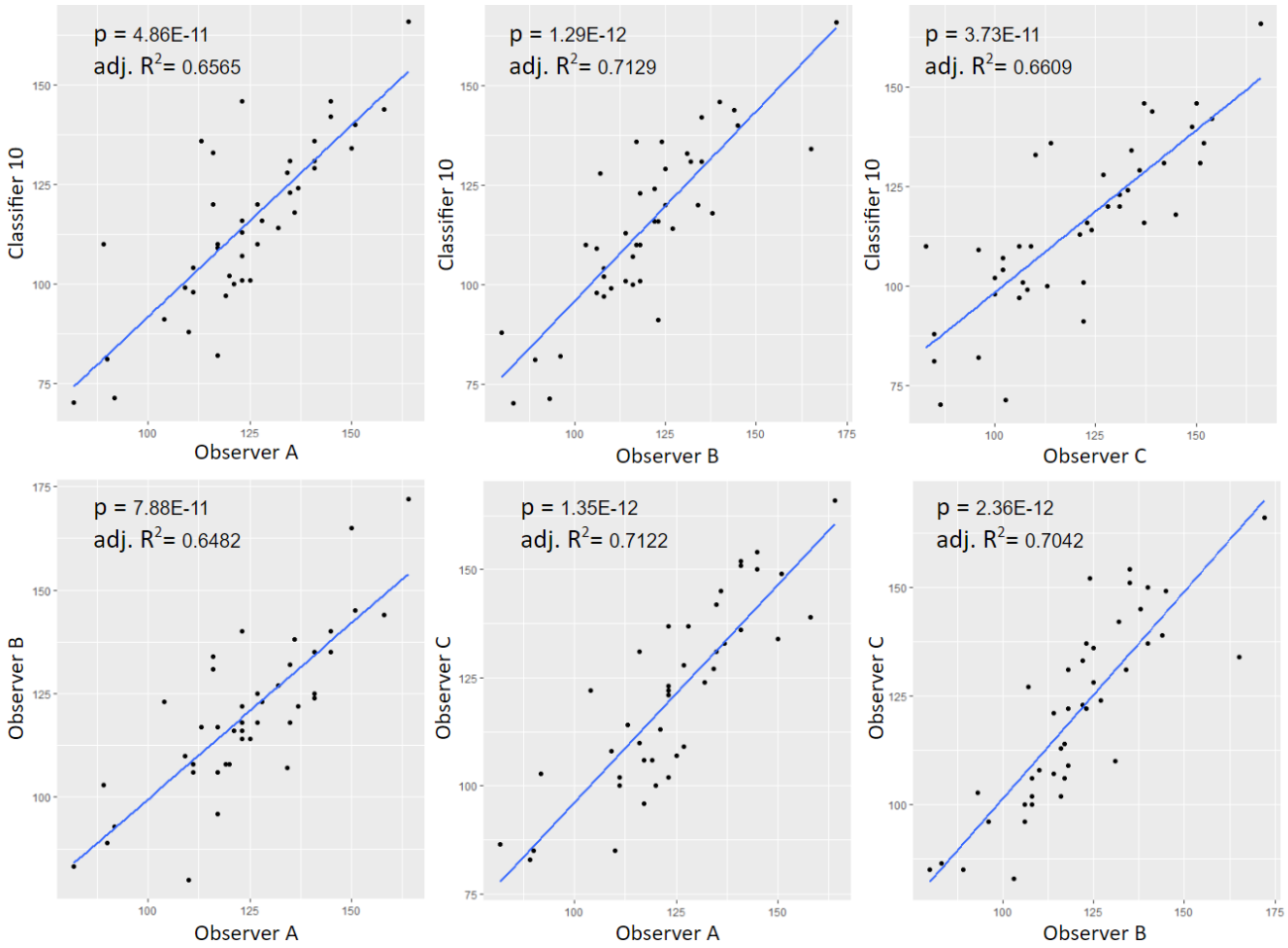


Figure S3. Scatter plots with repression lines for all correlative comparisons between observer and automatic counts. Automatic counts from classifier 10 for the experimental dataset used for automatic counts. All relationships showed significant overall correlations (p and Adj. R^2 values included in figure; n = 42).

Figure S4. ACCT applied on image of aerial photography. An image taken from a public dataset by Yan et al. with permission granted under MIT license. The image consists of houses in a field along with the image segmented for houses by ACCT. a threshold of 0.5 was used with a minimum pixel size for objects of 300 and maximum of 1000. The segmentation shows the rough positions or irregularities in the landscape, demonstrating the potential of ACCT outside the biological context.

Supplementary_Table_T1. Best 10 Classifier's Performance in the Validation Stage of the Iba-1 Dataset						
classifier	precision	recall	F1	accuracy	MAE	MPE
classifier4	0.948567	0.815217	0.876851	0.780708	19.4	0.014058
classifier2	0.944915	0.807971	0.871094	0.771626	20	0.014493
classifier10	0.971243	0.758696	0.851912	0.742027	30.2	0.021884
classifier5	0.968606	0.760145	0.851807	0.741867	29.7	0.021522
classifier9	0.971857	0.750725	0.847097	0.734752	31.4	0.022754
classifier8	0.970037	0.750725	0.846405	0.733711	31.2	0.022609
classifier21	0.976099	0.739855	0.841715	0.72669	33.4	0.024203
classifier11	0.972407	0.74058	0.840806	0.725337	32.9	0.023841
classifier7	0.972355	0.73913	0.839852	0.723918	33.1	0.023986
classifier16	0.97541	0.689855	0.808149	0.678063	40.4	0.029275
classifier17	0.977249	0.684783	0.805283	0.674037	41.3	0.029928
classifier1	0.97319	0.526087	0.682973	0.518571	63.4	0.045942
classifier3	0.968571	0.245652	0.391908	0.24371	103	0.074638

Supplementary Table T1. Top 10 classifiers on Iba-1 validation data. Classifiers are trained iteratively in TWS using small additions in training data. Reported statistics are calculated from the total of all cells in each image in the dataset except for mean absolute error (MAE) and mean paired error (MPE) (n = 10 images). Statistics include the following categories. Precision which is the number of true positive automatically counted cells divided by the total number of automatically counted cells. Recall which is the number of true positive automatically counted cells divided by the total number of hand placed counts. F1 which is the harmonic average of precision and recall. Accuracy which is the number of true positive automatically counted cells divided by the total of the number of automatically counted cells plus false negatives.

Label	Area	Min	Max	X	Y	XM	YM	Perim.	BX	BY	Width	Height
37_y.tif:0001-0409	814	0.094126	0.309577	1194.794	409.2666	1193.983	409.4465	156.066	1175	395	42	29
37_y.tif:0002-0429	423	0.09723	0.280072	1209.311	430.8097	1208.298	431.0496	96.91169	1195	419	28	21
37_y.tif:0003-0458	931	0.092795	0.249532	1221.167	456.877	1220.868	456.5184	198.6934	1195	441	51	35
37_y.tif:0004-0481	315	0.097334	0.190162	924.4111	479.7571	924.3771	479.7212	91.29646	912	470	24	22
37_y.tif:0005-0524	267	0.093953	0.155944	849.0805	522.2753	849.0322	522.3671	80.3259	840	512	18	25
37_y.tif:0006-0550	520	0.100025	0.293315	832.9827	548.3808	832.7923	547.9538	125.6812	823	529	24	42
37_y.tif:0007-0583	787	0.090755	0.292631	1383.156	586.702	1382.556	587.6921	187.865	1362	564	38	38
37_y.tif:0008-0629	253	0.092346	0.275275	1186.385	629.6146	1185.99	629.451	74.42641	1174	622	28	14
37_y.tif:0009-0676	858	0.095693	0.286489	1508.628	678.1364	1508.544	678.9255	184.7939	1492	656	34	40

Supplementary table T2. This is an example of the per cell morphological measurements recorded by ACCT on a single image in the Fluocell dataset. The morphological results are calculated for the following variables and measures described at https://imagej.nih.gov/ij/docs/menus/analyze.html. Label is the identification number of a cell counted by ACCT in the listed image file; formatted as the file name, identification number of the cell. Size and location measurements in pixels.

Conclusion

The immune landscape of the brain is a delicate battlefield in constant flux: immunity must come from wide populations of cells working together behind the blood-brain barrier and beyond the reach of general adaptive immunity (Kanmogne et al., 2020; Kaul et al., 2001). Here, microglia are relied on to quickly dispose of infected or malfunctioning cells in a tightly controlled manner to avoid damage to the precarious communication channels between neurons (Laurence, 1993).

Changing cell populations make it difficult to resolve the progression of disease from measures of complete gene expression (Capurro et al., 2014). Here, I employed linear modeling concepts in a rapid computational setting to uncover previously hidden trends in gene expression (Kuhn et al., 2011). These relationships implicated regulatory elements of some of the central pathways involved in viral immune response in the brain, i.e. the interferon response. Further, this particular set of pathways has been identified as a driving force in HIV induced neurodegeneration.

Previous studies demonstrated that GBP family members played diverse roles in responses to pathogens via innate immunity (Feng & Man, 2020; Kim et al., 2011; Shenoy et al., 2012). We found that Gbp3/GBP4 were upregulated in conjunction with immune cell populations through Population Specific Enrichment Analysis of whole brain microarray expression data, as well as RNA Seq in cultured human immune cells. The PSEA analysis both identified a positive correlation between putative microglia population size in the brain samples and Gbp3, and an

increase in Gbp3 in HIV model animals when accounting for that existing positive relationship. This analysis involved bulk linear model creation and filtering for genes where the best fitting model only involved a single cell population measure. Gbp3 was one of the genes quantitatively most highly associated with a single cell type, namely microglia. The question of whether or not Gbp3 is in fact localized to microglia or other cell types in the brain remains to be addressed experimentally, for example using immunofluorescence staining and quantification using ACCT.

Further RNA expression level information indicated an increase in GBP4 and IRF7 expression with HIV infection in human monocyte derived macrophage cell culture. The RNA sequencing also revealed a negative trend in expression between GBP4 and IRF7. IRF7 is an important promoter of the interferon pathways (Sun & Wang, 2012). This interaction could result in control of the chronic immune activation resulting from HIV related chronic immune activation in the brain.

Examining these genes in human cell culture revealed a positive, but muted response of GBP3 and IRF7 during immune activation prompted by exposure to the bacterial membrane lipid-bound sugar, LPS. We compared the expression of GBP4 and IRF7 to GBP1 and IFIT1, genes canonically activated by the type II interferon response which has been identified previously to respond to bacterial threats. GBP1 and IFIT1 were more significantly upregulated, indicating that the LPS stimulation was successful. However, the brain-resident microglia may not possess the same levels of reactivity as the monocyte-derived macrophages for the study of our genes of interest in the interferon pathways (Dello Russo et al., 2018).

Preliminary methodology analysis for determining Gbp3 colocalization revealed interesting avenues for further research. Gbp3 itself was widespread, but appears highly variable in an as-of-yet unidentified cell type. In attempting to verify existing microglial markers, I observed full separation of cell body staining for Iba-1 and P2ry12 in mouse cortex in both wild type animals and our transgenic model of neuroHIV.

In the crucial work of understanding the pathways of HIV infection and the resulting immune response, digital information which may hold the key to understanding and reversing damage is largely processed slowly, by hand. While cell counting software has advanced rapidly, the cutting edge has left behind labs with less computational resources and experience (Goecks et al., 2010). So, I have developed ACCT: Automatic Cell Counting with Trainable Weka Segmentation.

This program requires no large external computational resources and minimal user experience. At no point is the user required to edit code files while the program applies numerous Python and ImageJ scripts to assisting in training, validation and implementing machine learning cell counting models. This program is freely available on Github, allowing users in neuroimmunology and beyond access to an alternative to other automatic cell quantification platforms that is unique in its accessibility and complete validation and auditing structure.

Further development for ACCT will progress in both effectiveness and accessibility. The next major advancement for the program will be enabling the use of 3D image data. This information provides much greater clarity of cellular localization, especially in tissue, but presents a logistical challenge in updating the methods of accuracy assessment to accommodate

an additional axis of variation. To increase accessibility, the next step will be to package the program and its scripts into the ImageJ GitHub repository directly for automatic updating and access bundled with installations of ImageJ itself.

Now that GBP4/Gbp3 have been implicated in HIV-induced neurodegeneration through interaction with the interferon pathways, further research can focus on confirming the downstream effects of the gene in neuroHIV. To examine this interaction in a human cell model, a coculture of neurons, astrocytes, and microglia can be used to simulate the intercellular interactions critical to immune response in the brain as was used in previous studies of HIV-induced neurodegeneration from the Kaul lab (Thaney et al., 2017). With exposure to HIV protein, we could determine if the localization observed in mouse brains of Gbp3 holds for the ortholog GBP4 in human cells. This would identify a central component of the central nervous system's immune response to HIV.

The goal of this dissertation has been to present a handhold for future researchers on the journey to greater understanding of the role of important components of the innate immune system, interferons and microglia, in NeuroHIV. I have provided context on the complex process of adverse immune response in the nervous system. GBP4 is positioned to play an important role in the innate immune response to HIV, and the trajectory towards damage as the response progresses. I have also presented a tool to accelerate difficult research with greater consistency. ACCT was designed with accessibility in mind, making the most of existing analysis tools with Trainable Weka Segmentation, and built up in ImageJ, a popular, free imaging platform already employed by many labs.

References

Capurro, A., Bodea, L.-G., Schaefer, P., Luthi-Carter, R., & Perreau, V. M. (2014). Computational deconvolution of genome wide expression data from Parkinson's and Huntington's disease brain tissues using population-specific expression analysis. *Frontiers in Neuroscience*, *8*, 441.

Dello Russo, C., Cappoli, N., Coletta, I., Mezzogori, D., Paciello, F., Pozzoli, G., Navarra, P., & Battaglia, A. (2018). The human microglial HMC3 cell line: where do we stand? A systematic literature review. *Journal of Neuroinflammation*, *15*(1), 259.

Feng, S., & Man, S. M. (2020). Captain GBP1: inflammasomes assemble, pyroptotic endgame. *Nature Immunology*, *21*(8), 829–830.

Goecks, J., Nekrutenko, A., Taylor, J., & Galaxy Team. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, *11*(8), R86.

Kanmogne, G. D., Fonsah, J. Y., Umlauf, A., Moul, J., Doh, R. F., Kengne, A. M., Tang, B., Tagny, C. T., Nchindap, E., Kenmogne, L., Franklin, D., Njamnshi, D. M., Kuate, C. T., Mbanya, D., Njamnshi, A. K., & Heaton, R. K. (2020). Effects of HIV infection, antiretroviral therapy, and immune status on the speed of information processing and complex motor functions in adult Cameroonians. *Scientific Reports*, *10*(1), 14016.

Kaul, M., Garden, G. A., & Lipton, S. A. (2001). Pathways to neuronal injury and apoptosis in HIV-associated dementia. *Nature*, *410*(6831), 988–994.

Kim, B.-H., Shenoy, A. R., Kumar, P., Das, R., Tiwari, S., & MacMicking, J. D. (2011). A family of IFN-γ-inducible 65-kD GTPases protects against bacterial infection. *Science*, *332*(6030), 717–721.

Kuhn, A., Thu, D., Waldvogel, H. J., Faull, R. L. M., & Luthi-Carter, R. (2011). Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain. *Nature Methods*, *8*(11), 945–947.

Laurence, J. (1993). T-cell subsets in health, infectious disease, and idiopathic CD4+ T lymphocytopenia. *Annals of Internal Medicine*, *119*(1), 55–62.

Shenoy, A. R., Wellington, D. A., Kumar, P., Kassa, H., Booth, C. J., Cresswell, P., & MacMicking, J. D. (2012). GBP5 promotes NLRP3 inflammasome assembly and immunity in mammals. *Science*, *336*(6080), 481–485.

Sun, B., & Wang, J. (2012). Guanylate binding protein (GBP) 4 negatively regulates virus induced type I interferon and antiviral response by targeting interferon regulatory factor (IRF) 7 (168.30). *The Journal of Immunology*, *188*(1 Supplement), 168.30–168.30.

Thaney, V. E., O'Neill, A. M., Hoefer, M. M., Maung, R., Sanchez, A. B., & Kaul, M. (2017). IFNβ Protects Neurons from Damage in a Murine Model of HIV-1 Associated Brain Injury. *Scientific Reports*, *7*, 46514.