

2 Years of Open Notebooking; Lessons Learnt from Labscribbles

Dr Rachel Harding, University of Toronto Structural Genomics Consortium

Open Lab Notebooks: An Extreme Open Science Initiative CIHR Ottawa, Canada Friday 19th January 2018



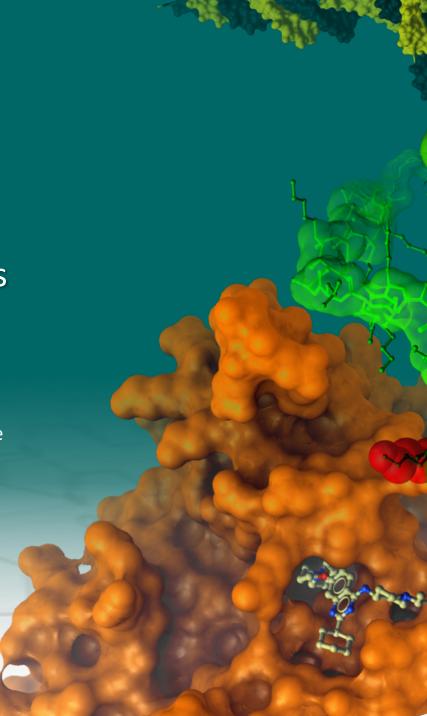






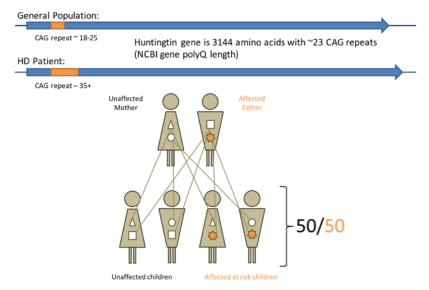




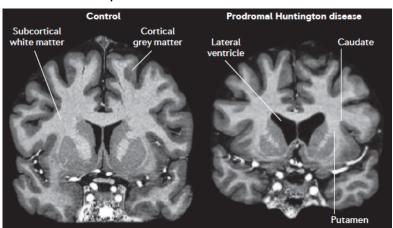


Huntington's disease (HD) is a progressive neurodegenerative disorder

HD patients have a CAG expansion in the 5' of the HTT gene. HTT mutation is inherited in an autosomal dominant fashion

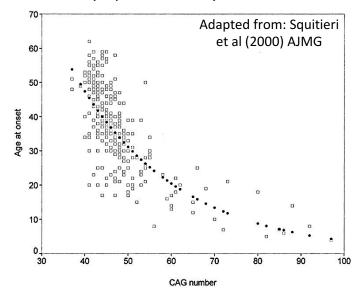


Progressive neurodegenerative disorder with a wide range of symptoms observed above a critical threshold of 35-40 CAG repeats



Adapted from: Bates et al (2015) Nature Disease Primers

CAG repeat length correlates with symptom age of onset and symptom severity



Movement disorders

- Involuntary jerking or writhing movements (chorea)
- Muscle problems, such as rigidity or muscle contracture (dystonia)
- Slow or abnormal eye movements
- Impaired gait, posture and balance
- Difficulty with the physical production of speech or swallowing
- Impairments in voluntary movements — rather than the involuntary movements — may have a greater impact on a person's ability to work, perform daily activities, communicate

and remain independent.

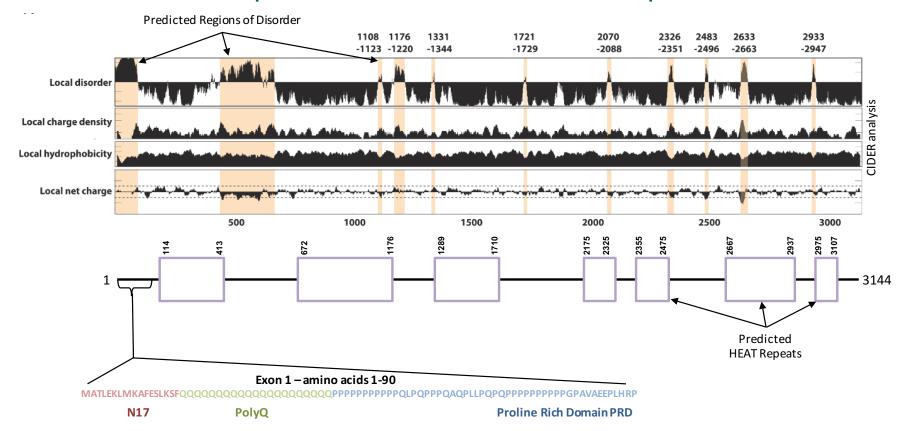
Cognitive disorders

- Difficulty organizing, prioritizing or focusing on tasks
- Lack of flexibility or the tendency to get stuck on a thought, behavior or action (perseveration)
- Lack of impulse control that can result in outbursts, acting without thinking and sexual promiscuity
- Lack of awareness of one's own behaviors and abilities
- Slowness in processing thoughts or "finding" words
- Difficulty in learning new information

Psychiatric disorders

- DepressionFeelings of irritability, sadness or apathy
- Social withdrawal
- Insomnia
- Fatigue and loss of energy
- Frequent thoughts of death, dying or suicide
- Other common psychiatric disorders include:
- Obsessive-compulsive disorder, a condition marked by recurrent, intrusive thoughts and repetitive behaviors
- Mania, which can cause elevated mood, over-activity, impulsive behavior and inflated self-esteem
- Bipolar disorder, or alternating episodes of depression and mania
- In addition to the above symptoms, weight loss is common in people with Huntington's disease, especially as the disease progresses.

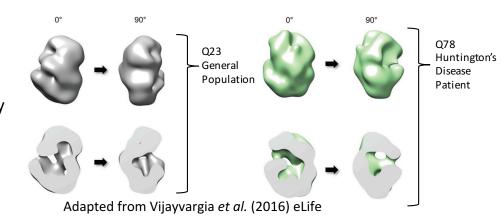
Why HTT with 40+ repeats is devastatingly damaging but benign with less than 35 repeats remains a fundamental question in HD



Huntingtin (HTT) is through to be a large scaffold protein - 100+ binary HTT:protein interactions reported

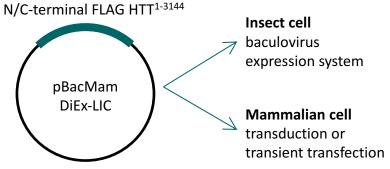
Unstructured regions of HTT are heavily post-translationally modified

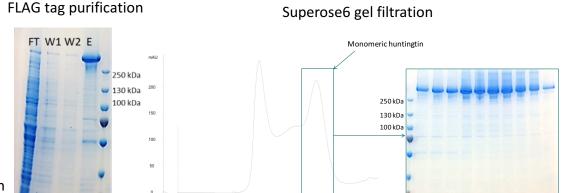
Low resolution EM reveals a hollow structure but with no significant global changes upon polyglutamine expansion



HD project aim: characterise HTT protein structure and binding partners

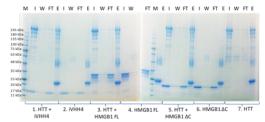
1. Purify full-length HTT protein



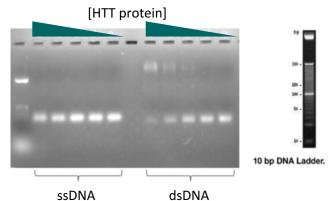


2. Identify interaction partners

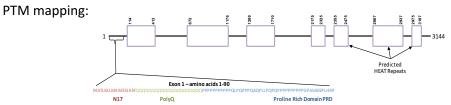
Protein:protein e.g. HMGB1, PRC2, nanobody



Protein:ligand e.g. DNA:



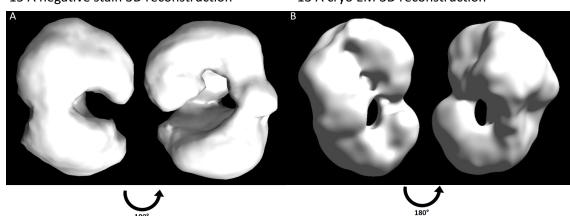
3. Biophysical characterisation and structural biology of HTT



Electron microscopy:

15 Å negative stain 3D reconstruction

13 Å cryo EM 3D reconstruction



Progress towards understanding HD and finding therapies is slow



A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntington's Disease Chromosomes

The Huntington's Disease Collaborative Research Group*

Summary

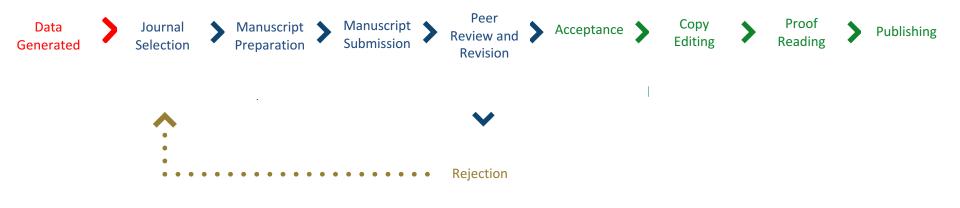
The Huntington's disease (HD) gene has been mapped in 4p16.3 but has eluded identification. We have used haplotype analysis of linkage disequilibrium to spotlight a small segment of 4p16.3 as the likely location of the defect. A new gene, IT15, isolated using cloned trapped exons from the target area contains a polymorphic trinucleotide repeat that is expanded and unstable on HD chromosomes. A (CAG), repeat longer than the normal range was observed on HD chromosomes from all 75 disease families examined, comprising a variety of ethnic backgrounds and 4p16.3 haplotypes. The (CAG), repeat appears to be located within the coding sequence of a predicted ~348 kd protein that is widely expressed but unrelated to any known gene. Thus, the HD mutation involves an unstable DNA segment, similar to those described in fragile X syndrome, spino-bulbar muscular atrophy, and myotonic dystrophy, acting in the context of a novel 4p16.3 gene to produce a dominant phenotype.

Introduction

Huntington's disease (HD) is a progressive neurodegenerative disorder characterized by motor disturbance, cognitive loss, and psychiatric manifestations (Martin and Gusella, 1986). It is inherited in an autosomal dominant fashion and affects ~ 1 in 10,000 individuals in most populations of European origin (Harper et al., 1991). The hallmark of HD is a distinctive choreic movement disorder that typically has a subtle, insidious onset in the fourth to fifth decade of life and gradually worsens over a course of 10 to 20 years until death. Occasionally, HD is expressed in juveniles, typically manifesting with more severe symptoms including rigidity and a more rapid course. Juvenile onset of HD is associated with a preponderance of paternal transmission of the disease allele. The neuropathology of HD also displays a distinctive pattern, with selective loss of neurons that is most severe in the caudate and putamen. The biochemical basis for neuronal death in HD has not yet been explained, and there is consequently no treatment effective in delaying or preventing the onset and progression of this devastating disorder.

The genetic defect causing HD was assigned to chromosome 4 in 1983 in one of the first successful linkage analyses using polymorphic DNA markers in humans (Gusella

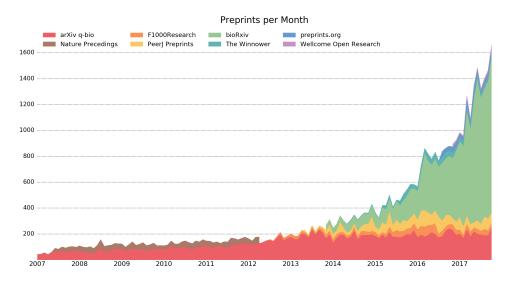
Flaws in the traditional publishing workflow are highlighted increasingly



This model has some issues...

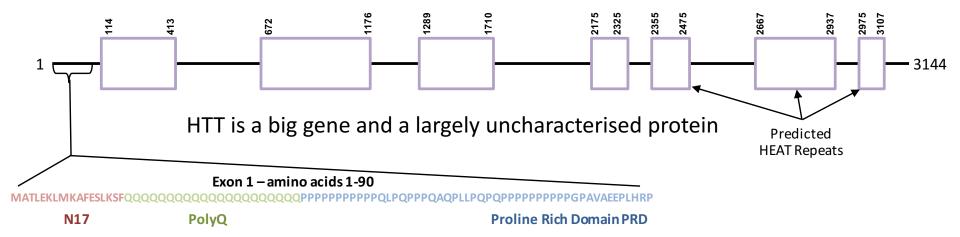
- 1. Publication bias
- 2. The process is slow
- 3. Its expensive
- 4. Peer review is problematic
- 5. Often lacking data/details
- 6. Copyright issues
- 7. Often paywalled

...and preprints solve some of these issues



Jordan Anaya http://asapbio.org/preprint-info/biology-preprints-over-time

HD research tools are difficult to generate and it's a challenging target



What type of research might not get published?

- "Failed" projects no big, shiny, fashionable conclusions
- Difficult methodological issues to resolve
- Intermediate results incomplete research project due to staff departure, shift in focus, grant lapse
- Contradict published study proving a negative is more difficult that a positive

The argument for open access release of this data:

- Its been paid for with tax dollars transparent, complete and honest reporting of projects
- There are useful nuggets of information (data and protocols)
 which can help other researchers with similar projects
- Review by peers may help solve problems to move the work forward



Open Notebook Research

Open access research

Collaborative and interactive

Real-time and honest reporting

Accessible to all

Points of consideration for sharing data

Quality – excellent science, high level curation

Detail – materials and methods, data entry, easy reuse

Speed – fast bench to public domain timeframe

Outreach – data communicated to researchers and wider spheres

Presentation – clear and self-explanatory

Feedback – dialogue with readers

Discoverability – easy to find

Impact – researchers using and building on data

Best practises for open notebooking - inspiration from others

ACTIVE:

Real-Time Reports, Truant lab - McMaster (Neurodegeneration) https://raytruantlab.wordpress.com ACTIVE

Open Source Malaria (Chemistry) http://opensourcemalaria.org ACTIVE

Rosanne Hertzberger (Microbiology) http://www.reblab.org ACTIVE

Carl Boettinger (Ecology) http://www.carlboettiger.info/lab-notebook.html ACTIVE

Dror Batan (Mathematics) http://drorbn.net/AcademicPensieve/2017-11/index.html ACTIVE

Open Notebook Science Network (ONSN) http://onsnetwork.org Catalog - ACTIVE

Tobias Osborne (Theoretical Physics) https://tjoresearchnotes.wordpress.com Teaching resource - ACTIVE

Andres G. Saravia https://research-engine.appspot.com/andres Archive - ACTIVE

INACTIVE:

Scott Veirs (Marine Biology) http://onsnetwork.org/scottveirs/ INACTIVE

Ann Mayo (Ecology) http://onsnetwork.org/mayonotebook/category/research/ INACTIVE

Chem-bla-ics (Computational Chemistry) http://onsnetwork.org/chemblaics/ INACTIVE

Nickolas J. Lasorte https://nicklasorte.wordpress.com INACTIVE

Anthony Salvagno (Open Access Policy) http://research.iheartanthony.com INACTIVE

Jean-Claude Bradley (Chemistry)

http://usefulchem.wikispaces.com/?responseToken=0ad5419ef05b77998507275d4009610b8 INACTIVE

Brigitte Black (Physics) https://openwetware.org/wiki/User:Brigette_D._Black/Notebook/Brigettes_Notebook INACTIVE

Jeremiah Faith (Molecular Biology)

http://www.jeremiahfaith.com/open notebook science/jeremiah faith lab notebook.pdf INACTIVE

Nadiez Fernandez-Oropeza (Biochemistry) https://openwetware.org/wiki/User:Nadiezda_Fernandez-

Oropeza/Notebook/Notebook

Andy Maloney (Microscopy) https://andymaloney.wordpress.com INACTIVE

Cameron Neylon (Biophysics) http://biolab.isis.rl.ac.uk/projects/blog/blogs.php/blog_id/5 INACTIVE

Andrew Rambaut (Epidemiology) http://influenza.bio.ed.ac.uk INACTIVE

NB: Defined as inactive if no posts in last 6 months. Numerous examples software/code project open notebooks on GitHub.

Best practises for open notebooking - definition of a notebook vs. blog

Open Lab Notebook:

- Making complete research project available online as it is happens
- Compiled of project plans, experimental protocols and setups, raw data, unfiltered analyses and other relevant works
- Created and run on a platform to suit scientists needs but available to anyone to read

Lab Blog:

- Selective in reporting research outcomes
- Generalised updates lacking experimental details or to include raw data

Best practises for open notebooking - guidelines for Labscribbles:

 Every aspect of this project, conducted by myself, will be written up for the notebook

- 2. The level of detail and quality of write up will be equivalent to the electronic lab notebook system (existing infrastructure at SGC)
- 3. Experiments will be written up as soon as possible upon completion
- 4. Maintain fluidity in structure and format for the project to evolve

The format of my open notebook





Experiments in the lab, commentary on scientific meetings and published reports

Written in report on labscribbles.com

Link to scientific findings deposited in Zenodo under creative commons licensing

Share information on social media platforms





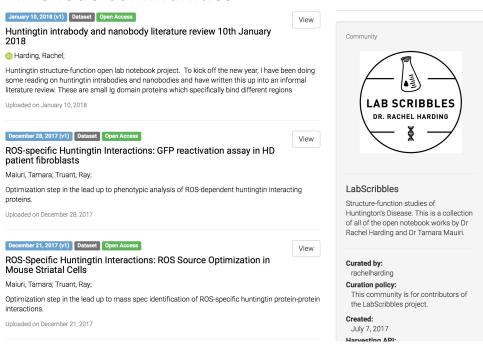






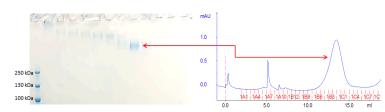
The format of my open notebook - changes made since its inception

1. Zenodo communities



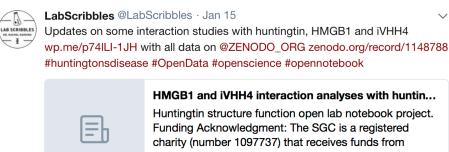
2. Images on blog

We were able to replicate Ihn Sik's expression and purification of the full-length huntingtin, first in a small-scale trial experiment. Next I worked to try and optimise the purification protocol to make a quality protein sample which we can then use for analysis and further experiments. The first few attempts did not work very well – the protein was very aggregated meaning that all of the protein molecules were sticking to each other which is not ideal for subsequent experiments. However, trying a few different conditions to see which might help purify the huntingtin protein more stringently, I landed on a set of conditions which work really well and make a sample which looks similar to that which Ihn Sik describes in his paper. Since then I have optimised a procedure called gradient fixation (Grafix) which chemically fixes the huntingtin protein to make it more stable, possibly a critical step in generating a sample for structural analyses.



Huntingtin protein fractions from the grafix procedure were confirmed to be stable monomeric huntingtin by analysis on a gel filtration column.

3. Tweeting links to data and blog



zenodo.org

4. Cutting time burden and duplication of effort Write up at end of experiment completion Notebook write up -> cut and paste for weekly meetings Succinct blog posts but link to previous/relevant steps for full story

The format of my open notebook - changes made since its inception

5. File format in uploads and annotation

FAIR Data Principles:

- -> Findable
- -> Accessible
- -> Interoperable
- -> Re-usable

Self-explanatory titles and PDF preview:

Check points:

- -> Non-proprietary format
- -> Easy to extract and reuse data/info
- -> Discoverable
- -> Makes sense?

HMGB1 and iVHH4 interaction analyses with huntingtin by pull-down - 15th January 2018

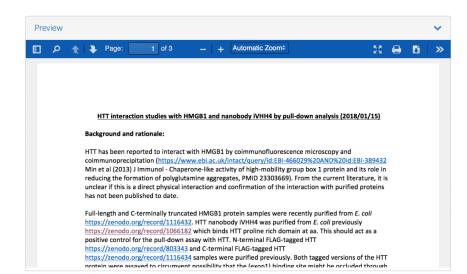
Dataset Open Access

(b) Harding, Rachel; Arrowsmith, Cheryl; Edwards, Aled

January 15, 2018

Huntingtin structure function open lab notebook project

Funding Acknowledgment: The SGC is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for Innovation, Eshelman Institute for Innovation, Genome Canada through Ontario Genomics Institute [OGI-055], Innovative Medicines Initiative (EU/EFPIA) [ULTRA-DD grant no. 115766], Janssen, Merck KGaA, Darmstadt, Germany, MSD, Novartis Pharma AG, Ontario Ministry of Research, Innovation and Science (MRIS), Pfizer, São Paulo Research Foundation-FAPESP, Takeda, and Wellcome.



Content of the notebook

Literature reviews

January 10, 2018 (v1) Dataset Open Access

View

Huntingtin intrabody and nanobody literature review 10th January 2018

Harding, Rachel;

Huntingtin structure-function open lab notebook project. To kick off the new year, I have been doing some reading on huntingtin intrabodies and nanobodies and have written this up into an informal literature review. These are small Ig domain proteins which specifically bind different regions

Uploaded on January 10, 2018

Data analyses

November 6, 2017 (v1) Dataset Open Access

Mapping of phosphorylation modifications of insect cell derived huntingtin (2017/11/06)

Harding, Rachel; Ackloo, Suzanne; Arrowsmith, Cheryl; Edwards, Aled;

Huntingtin structure function open lab notebook

Uploaded on November 6, 2017

Computational analysis

View

March 29, 2016 (v1) Dataset Open Access

RaptorX structure prediction of huntingtin

Harding, Rachel; Toledo-Sherman, Leticia; Arrowsmith, Cheryl; Edwards, Aled;

Huntingtin structure-function open lab notebook project.

Uploaded on March 29, 2016

Experimental plans

March 24, 2016 (v1) Dataset Open Access

Huntingtin domain construct design for insect cell expression

Harding, Rachel; Toledo-Sherman, Leticia; Arrowsmith, Cheryl; Edwards, Aled;

Huntingtin structure-function open lab notebook project

Uploaded on March 29, 2016

Conference presentation

View

April 26, 2017 (v1) Poster Open Access

CHDI Conference 2017 Poster: Pursuit of a high resolution structure of fulllength huntingtin by cryo-electron microscopy

Harding, Rachel; Deme, Justin; Loppnau, Peter; Ackloo, Suzanne; Hutchinson, Ashley; Hunt, Brittany; Seitova, Alma; Lea, Susan; Arrowsmith, Cheryl; Edwards, Aled; Holehouse, Alex;

Huntington's disease is hallmarked by the CAG expansion of the huntingtin gene. How the corresponding polyQ expansion affects the structure of the encoded huntingtin protein remains poorly understood in the absence of a high resolution full-length protein structure. Huntingtin is a large, monomeric

Uploaded on April 26, 2017

View

View

View

January 15, 2018 (v1) Dataset Open Access

Wet-lab experiment write up

HMGB1 and iVHH4 interaction analyses with huntingtin by pull-down - 15th January 2018

(i) Harding, Rachel; Arrowsmith, Cheryl; Edwards, Aled;

Huntingtin structure function open lab notebook project. Funding Acknowledgment: The SGC is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for Innovation, Eshelman Institute for Innovation, Genome Canada through

Uploaded on January 15, 2018

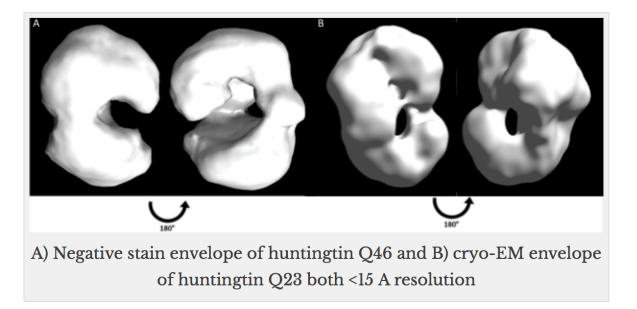
Content of the notebook - the good and the not so good

Cryo-EM updates and more!

■ August 11, 2017 👗 racheljaneharding 🗭 Leave a comment 🖉 Edit

This past few weeks have been busy with fellowship applications, writing grants and papers so I am rather behind on keeping you updated but there has been some exciting progress:

1. Susan Lea has made fantastic progress processing cryo-EM data of our samples and we now have a 13 A envelope of the protein: https://zenodo.org/record/841824



2. On the basis of Susan's work, I have again modified the sample preparation protocol and sent her some EVEN BETTER samples of huntingtin, hopefully bound to DNA. https://zeno-do.org/record/841878

Content of the notebook - the good and the not so good

Purification of high mobility group box 1 protein (HMGB1) – a huntingtin interacting protein

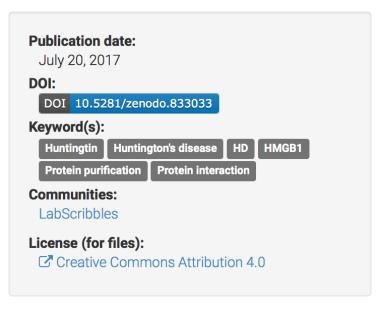
■ July 13, 2017 👗 racheljaneharding 🗭 Leave a comment 🖉 Edit

To generate better data for cryo-EM analysis of the huntingtin protein structure, more stable and conformationally homogeneous samples may help improve the particle averages. Purifying complexes of huntingtin with proteins which specifically bind huntingtin may promote such sample stability. High mobility group box 1 (HMGB1) is known to bind huntingtin but, to my knowledge, this interaction has not been demonstrated in the published literature with purified recombinant protein samples – just with confocal microscopy and co-immunoprecipitation studies. To begin to investigate this interaction, I first need to express and purify HMGB1.

This experiment was not very successful! I over-expressed HMGB1 no problem but the purification failed. However, I think I know how to fix it so am planning to repeat ASAP. You can read all of the details of how I did the experiments and how I plan to improve them on Zenodo.

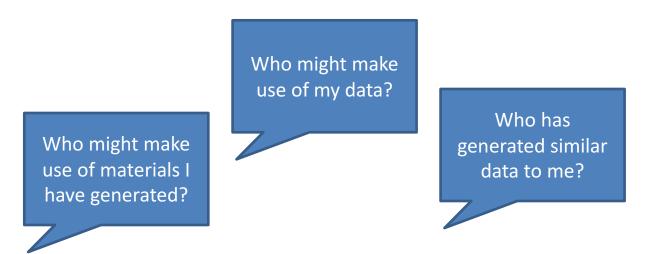
Discoverability of the notebook by different groups

1. Keywords in Zenodo uploads





3. Actively sharing the notebook or specific datasets



Caution! Open notebook status of data made clear in notebook posts and when shared with scientists directly

Working with collaborators on joint projects

1.	Every aspect of this project, conducted by myself, will be written up for the
	notebook

2. The level of detail and quality of write up will be equivalent to the electronic lab notebook system (existing infrastructure at SGC)

3. Experiments will be written up as soon as possible upon completion

4. Maintain fluidity in structure and format for the project to evolve

Working with collaborators on joint projects often means being open and flexible

1. Every aspect of this project, <u>conducted by myself</u>, will be written up for the notebook

2. The level of detail and quality of write up will be equivalent to the electronic lab notebook system (existing infrastructure at SGC)

- 3. Experiments will be written up as soon as possible upon completion
- 4. Maintain fluidity in structure and format for the project to evolve
- 5. All collaborators should be invited to be contributors of the open notebook and credited as authors on posts which they contribute to.....

Working with collaborators on joint projects often means being open and flexible

All collaborators should be invited to be contributors of the open notebook and credited as authors on posts which they contribute to.....

.....BUT collaboration opportunities should not be turned down if collaborators do not want work they conduct in the open notebook

January 10, 2017 Report Open Access

Collaborative Communication: Updates on cryoelectron microscopy experiments conducted by Prof. Susan Lea (2017/01/10)

Rachel Harding

Open lab notebook huntingtin structure function project.



Opportunities to travel and present my work to different audiences

Coordinating different project threads across multiple labs/collaborators

Improving my presenting skills and increasing my confidence in my own research

Running an opennotebook helps my ongoing professional development as an early career researcher

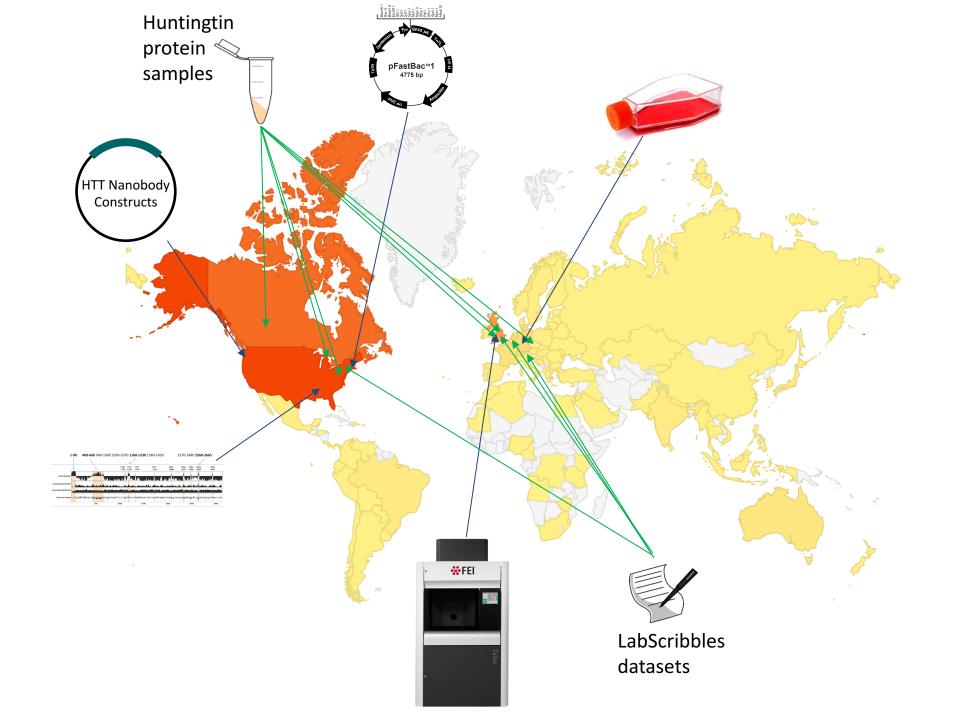
Project planning and milestone achievement

Raising my profile as an early career researcher

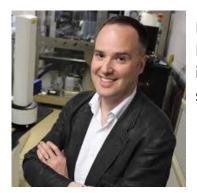
Honing writing skills – scientific and general audience







Open science helped me secure long-term collaborations and grants with field experts as well as mentorship:



Ray Truant – PI and HD researcher at McMaster. Working on HTT interaction partners and redox signalling.



Sarah Tabrizi – PI at UCL and head of the UCL HD research centre. Working together on analysis of HD patient SNPs



Susan Lea – PI at Oxford and manager of COSMIC EM facility. Working together to solve the structure of huntingtin by electron microscopy methods.



Bass Hassan – PI at Oxford, interested in transcription control. Would like huntingtin samples to probe interaction of the REST complex with huntingtin. Co-written successful grant awarded from Wellcome Trust.



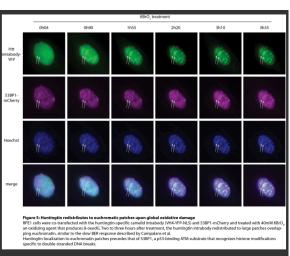
Jeff Carroll – PI and HD researcher at Western Washington University, Mentor and advisor and potential future collaboration.

HD researcher Dr. Tamara Maiuri started her own open notebook









VISUALIZING REAL HUNTINGTIN PROTEIN IN CELLS FROM AN HD PATIENT

Posted on October 31, 2017

Blog post by Dr. Tamara Maiuri

I am still busily collecting cells to be sent for mass spec for our goal of obtaining a list of proteins that interact with huntingtin upon oxidative DNA damage. Unfortunately I've run into a few road blocks, which I will blog about in the coming weeks (hopefully with a resolution!).

Meanwhile, I've been working on methods to assess the hit proteins for their physiological relevance as potential drug targets. Last time I described one such approach: the GFP reactivation assay. Since then, data from 3 experiments have been combined and look promising. While repair efficiency varies from experiment to experiment, mouse HD cells consistently show approximately half the repair efficiency of normal cells (an average of 44.8% over 3 experiments). This is a readout we can use to test the effects of manipulating our hit proteins.



ROS-specific Huntingtin Interactions: GFP reactivation assay in HD patient fibroblasts Maiuri, Tamara; Truant, Ray;

December 21, 2017 (v1) Dataset Open Access

 ${\hbox{ROS-Specific Huntingtin Interactions: ROS Source Optimization in Mouse Striatal Cells} \\$

October 31, 2017 (v1) Dataset Open Access

ROS-specific Huntingtin Interactions: Inducible huntingtin-specific chromobody expression by nucleofection

Maiuri, Tamara; Truant, Ray;

Maiuri, Tamara; Truant, Ray;

October 31, 2017 (v1) Dataset Open Access

ROS-specific Huntingtin Interactions: Comparing transfection methods for inducible expression of huntingtin-specific chromobody

Maiuri, Tamara; Truant, Ray;

October 3, 2017 (v1) Dataset Open Access

ROS-specific Huntingtin Interactions: G418 kill curve in TruHD fibroblasts

Maiuri, Tamara; Truant, Ray;

October 2, 2017 (v1) Dataset Open Access

ROS-specific Huntingtin Interactions: Testing inducible expression of huntingtinspecific chromobodies

Maiuri, Tamara; Truant, Ray;

October 2, 2017 (v1) Dataset Open Access

ROS-specific Huntingtin Interactions: Cloning huntingtin-specific chromobodies into pTRE-3G $\,$

Maiuri, Tamara; Truant, Ray;





LabScribbles

Structure-function studies of Huntington's Disease. This is a collection of all of the open notebook works by Dr Rachel Harding and Dr Tamara Mauiri.

Curated by:

rachelharding

Curation policy:

This community is for contributors of the LabScribbles project.

Created:

July 7, 2017

Harvesting API:

OAI-PMH Interface

Sharing my research online has given me opportunities to interact with HD patients and advocacy organisations



ACKNOWLEDGEMENTS



SGC and UHN

Peter Loppnau

Ashley Hutchinson

Brittany Hunt

Alma Seitova

Mani Ravichandran

Juliana The

Levon Halabelian

Suzanne Ackloo

Shili Duan

Cheryl Arrowsmith

Aled Edwards

Oxford University

Justin Deme

Susan Lea

Bass Hassan

Western Washington University

Jeff Carroll

UCL

Sarah Tabrizi

Ulm Univeristy

Stefan Kochanek

Harvard University

Ihn Sik Seong

McMaster University

Ray Truant Tam Mauiri

CHDI Foundation

Leticia Toledo-Sherman

Novartis

Christian Wiessmann

Sandra Jacob

Sick Kids Hospital

John Rubinstein



Accelerating therapeutic development for Huntington's disease

www.thesgc.org www.labscribbles.com @labscribbles rachel.harding@utoronto.ca

FUNDING PARTNERS

The SGC is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for Innovation, Eshelman Institute for Innovation, Genome Canada through Ontario Genomics Institute [OGI-055], Innovative Medicines Initiative (EU/EFPIA) [ULTRA-DD grant no. 115766], Janssen, Merck KGaA, Darmstadt, Germany, MSD, Novartis Pharma AG, Ontario Ministry of Research, Innovation and Science (MRIS), Pfizer, São Paulo Research Foundation-FAPESP, Takeda, and Wellcome [106169/ZZ14/Z].



Site: labscribbles.com

Email: rachel@labscribbles.com / rachel.harding@utoronto.ca

Twitter: @labscribbles