

FinnGen Sandbox Training: Breakout Rooms

Facilitator: Kumar Veerapen (veerapen@broadinstitute.org)

Date: December 8th, 2021

Instructions

1. Pick a someone to screen share
2. Pick a note taker who will
 - a. Create a live Google Doc / notepad / word doc for the group
 - b. Take notes
3. Pick one person to summarize what you did in 1-2 minutes
4. Everyone to go through the Google Doc action items
5. Smile and have a great time (videos on, hopefully)
6. Kumar will jump in and out of breakout rooms to take questions that you will have during the practice. If he doesn't know the answer, he will get back to you after the session! We always learn together :)

Use this:

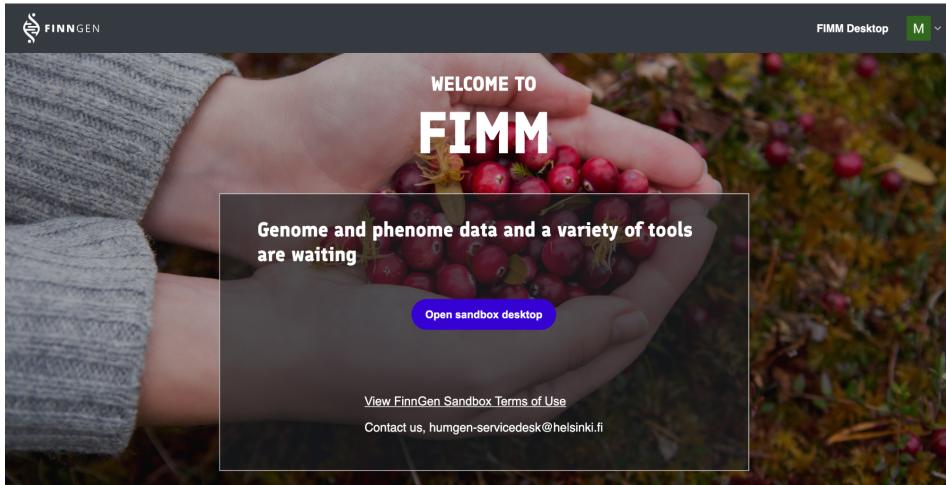
- Analyst Handbook: <https://app.gitbook.com/@finngen/s/finngen-analyst-handbook/>

Notes:

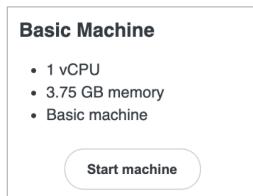
- 1) for this entire activity, you will need to have Sandbox access AND be configured for 2FA
- 2) You need to have your @FinnGen.fi ID ready

Activity 1: Logging in

1. Go to: <https://Sandbox.finngen.fi>
2. Log in using your @finngen.fi ID
3. Click on “Open sandbox desktop”

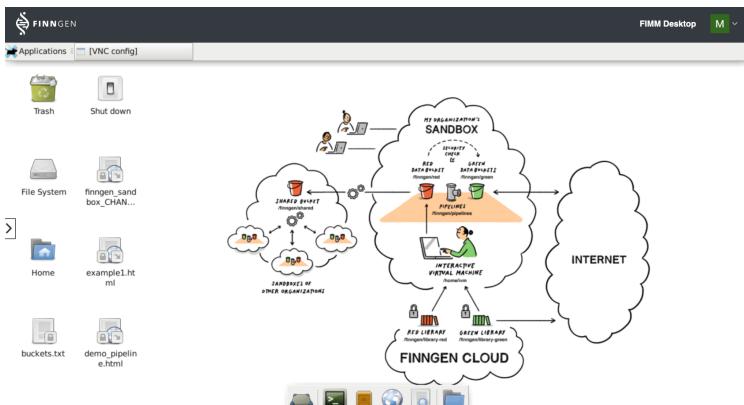


4. Choose “Basic Machine” by clicking “Start machine”



Activity 2: What are the available FinnGen Sandbox applications?

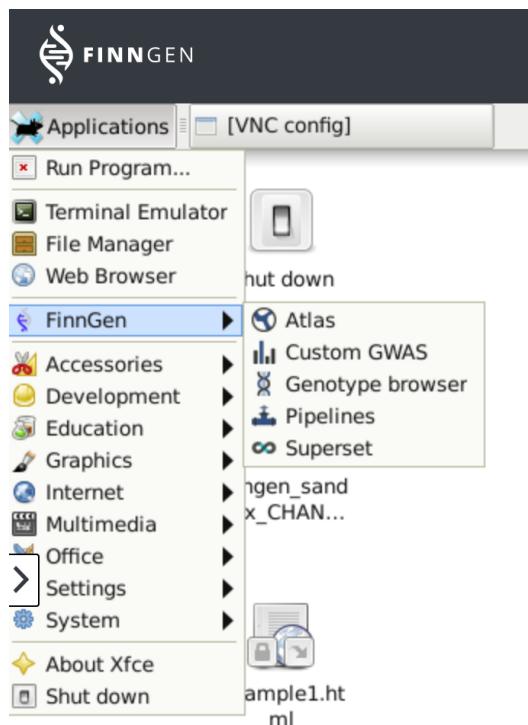
1. Take a look at your landing page and write down what you can see there to your own live document / notepad / word doc. The landing page should look like this:



Notice that this looks like any other desktop front i.e. Windows, MacOS except this is a Unix environment. You are now in your own interactive virtual machine using Unix as the operating system.

Also notice that there are icons where one of which was a document that we talked about earlier called `buckets.txt`. You do not need to open it now, but you're welcome to do it later on your own time.

2. Explore the environment for approximately 10-20 seconds by double clicking on
 - a. Home
 - b. File System
3. Let's explore the applications that are preloaded into your Sandbox environment



Each application is there for your use! More details can be found [here](#). To note, all tools are browser based and will open in a Mozilla Firefox browser.

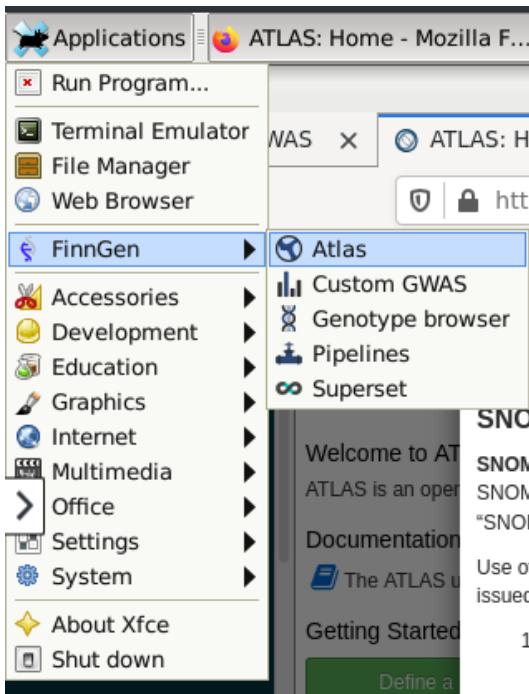
Activity 3: Custom GWAS

Yes, you read that right. Your very own custom GWAS without actually needing to code. Just a point and click away to creating your own phenotype and running a GWAS in minutes!

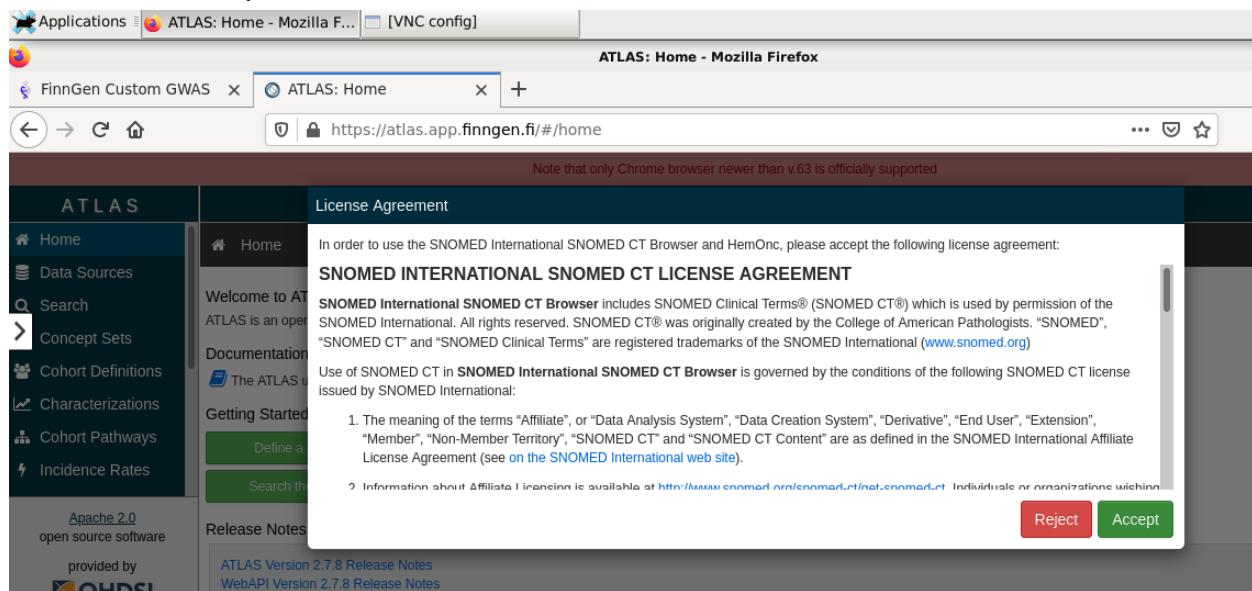
The materials for this activity were adapted from [here](#)

Activity 3.1: Defining your case-control cohort in Atlas

1. Click on Applications → FinnGen → Atlas



2. Click "Accept"



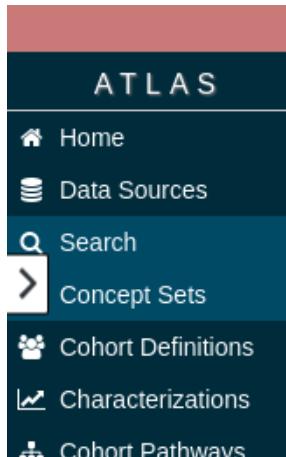
3. Select “Concept Sets” on the left hand bar
4. On the right, press the blue button that says “New Concept Set”

Concept Sets				
List		Export		
New Concept Set				
Show	10	entries		
Id	Title	Created	Modified	
358	Croup_non_standard[MN]	07/01/2021 10:58 AM	10/05/2021 8:27 AM	
464	LC_NE_Ind [VL]	10/05/2021 7:47 AM	10/05/2021 7:48 AM	
462	LC_97.2 [VL]	10/04/2021 12:49 PM	10/04/2021 1:55 PM	
461	ADPKD	10/01/2021 8:23 PM	10/01/2021 8:23 PM	
460	Test - adele	10/01/2021 7:52 PM	10/01/2021 7:52 PM	

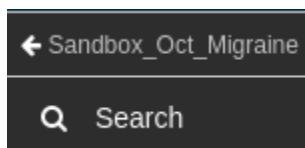
5. Go to the top and change where it says “*New Concept Set*” to be the name of your Concept set. I would suggest always starting with `Sandbox_month_migraine_groupNumber` (month for the month we are in, whatever phenotype to substitute `migraine` and `groupNumber` is your breakout room number). You are HIGHLY encouraged to create your own phenotype and define it in a bit. For the moment, we are just creating the phenotype name.

ACTION ITEM: *What is your phenotype of choice?*

6. By naming the concept set you could think that you have taken an empty shopping cart. Now you will go to the “Search” on the left side and add things.



7. In the black section near the top you will be able to see what *Concept Set*/shopping cart you are adding to.



8. In the “Search” menu you can try multiple spellings and combinations.

Notes:

RC = Record Count which is usually different from the number of individuals. You can sort by RC by clicking at the top of the column.

DRS = Descendants Records Count.

In this case, we are searching for migraine, so I used “migraine” as my keywords

The screenshot shows a search interface with a sidebar containing 'Vocabulary' and 'Class' sections. The main area displays a table of search results for 'migraine' with columns for Id, Code, Name, Class, RC, DRS, Domain, and Vocabulary. The results include entries like 'ANTIMIGRAINE PREPARATIONS' (ATC 3rd, 0, 974,493, Drug, ATC), 'Migraine' (Clinical Finding, 24,630, 60,591, Condition, SNOMED), 'Migraine with aura' (Clinical Finding, 16,059, 16,059, Condition, SNOMED), and 'Migraine with aura [classical migraine]' (ICD10f Hierarchy, 16,059, 16,059, Condition, ICD10f).

	Id	Code	Name	Class	RC	DRS	Domain	Vocabulary
	21604364	N02C	ANTIMIGRAINE PREPARATIONS	ATC 3rd	0	974,493	Drug	ATC
	318736	37796009	Migraine	Clinical Finding	24,630	60,591	Condition	SNOMED
	381549	4473006	Migraine with aura	Clinical Finding	16,059	16,059	Condition	SNOMED
	900004987	G43.1#	Migraine with aura [classical migraine]	ICD10f Hierarchy	16,059	16,059	Condition	ICD10f

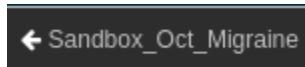
9. Select the ones you want to add to your Concept set by clicking on the *shopping cart icon* on the left of your desired phenotype.

In the Search some are in red – those are ones that are coming directly from the Finnish health registries medical codes (more about standard and non-standard codes [here](#)).

If you are using your phenotype of choice, click on the phenotype and see what the counts are before proceeding.

ACTION ITEM: *List your concept sets*

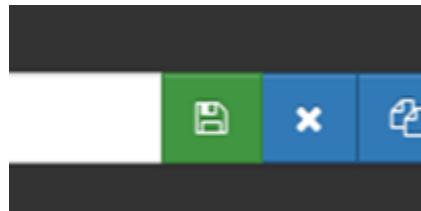
10. Once you have added all the codes click on the name of your concept set in the grey box on top.



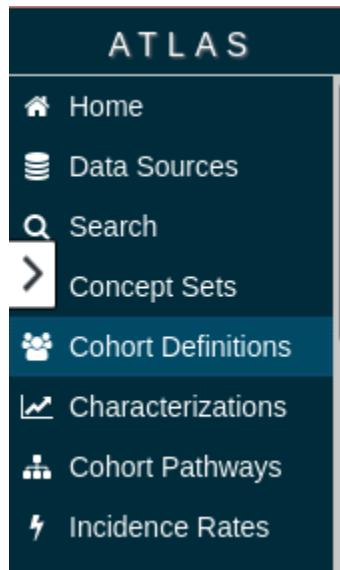
You will be redirected to a new screen.

The screenshot shows a software interface titled "Concept Set #466" with the sub-title "Sandbox_Oct_Migraine". The main area displays a table with one row of data. The columns are: Concept Id, Concept Code, Concept Name, Domain, Standard Concept Caption, Exclude, Descendants, and Mapped. The data row is: 318736, 37796009, Migraine, Condition, Standard, checked, unchecked, checked. Below the table are buttons for Classification (purple), Non-Standard (red), and Standard (blue). The top navigation bar includes tabs for "Concept Set Expression", "Included Concepts (1)", "Included Source Codes", "Explore Evidence", "Export", and "Compare". There are also buttons for "Optimize" and "Delete". A search bar and pagination controls (Show 25 entries, Previous 1 Next) are at the bottom.

11. Click save with the green disk icon.



12. Now go to “*Cohort Definitions*” on the left-hand panel.



13. Select “New Cohort”

The screenshot shows a table of cohort definitions with columns for Id, Name, Created, Updated, and Author. A message at the top right says "Please close your current cohort definition before creating a new one." The table includes entries like "Croup_non_standard[MN]", "Croup_<6years_IDC8-10[MN]", and "ADPKD controls [SVM]".

	Name	Created	Updated	Author
954	Croup_non_standard[MN]	10/05/2021 1:47 PM	10/05/2021 1:48 PM	anonymous
739	Croup_<6years_IDC8-10[MN]	07/01/2021 1:05 PM	10/05/2021 9:35 AM	anonymous
953	Croup_<6years_non_standard[MN]	10/05/2021 9:30 AM	10/05/2021 9:34 AM	anonymous
918	bimatoprost_control	09/21/2021 6:38 PM	10/05/2021 3:49 AM	anonymous
921	bimatoprost_case	09/22/2021 3:20 PM	10/05/2021 3:44 AM	anonymous
952	ADPKD controls [SVM]	10/01/2021 8:31 PM	10/01/2021 9:44 PM	anonymous

14. Fill in the text field at the top with your Cohort name. Here you will want to specify case and control cohorts in the name.

The screenshot shows the "New Cohort Definition" page with the "Definition" tab selected. It includes sections for "Cohort Entry Events" and "Inclusion Criteria". The "Cohort Entry Events" section has a note about continuous observation and initial events. The "Inclusion Criteria" section has icons for search, filters, and other database functions.

We will start with the case cohort and then duplicate it to add a control cohort.

For case cohort definition

15. Under the “Cohort Entry Events” section we define the rules dictating when patients are accepted into our cohort. Under the “Inclusion Criteria” section we define rules that further filter the patients accepted in the previous section.

For GWAS cohorts we are usually not interested in when the patients enter the cohort. Hence, for simplicity, in the Cohort Entry Events section, we are going to include all patients and, in the “Inclusion Criteria” section, we are going to filter these with a code in the concept set.

Under the “Cohort Entry Events” section, we include all patients at the first time that they got any visit in our database. We are just going to press “+Add Initial Event” and then scroll to the bottom of the picklist, and select “Add Visit” (you can choose a different event type based on your initial concept set)

The screenshot shows the ATLAS Cohort Definitions interface. On the left, there's a sidebar with links to Home, Data Sources, Search, Concept Sets, Cohort Definitions (which is selected), Characterizations, Cohort Pathways, and Incidence Rates. The main area has a header 'enter a cohort definition description here'. Below it is a 'Cohort Entry Events' section with a blue header. It contains fields for 'Events having any of the following criteria:' (with dropdowns for days before and after event index date), 'Limit initial events to:' (set to 'earliest event'), and a 'Restrict initial events' button. To the right of this is a sidebar with a 'Add Initial Event' button and a list of other selection methods: 'Add Condition Era', 'Add Condition Occurrence', 'Add Death', 'Add Device Exposure', 'Add Dose Era', and 'Add Visit'. At the bottom of the sidebar is a 'Delete' button.

16. After selection, it will look like this:

This screenshot shows the same 'Cohort Entry Events' section as the previous one, but with a single criterion selected: 'a visit occurrence of Any Visit'. The sidebar on the right is no longer present.

17. Now, let's define the inclusion criteria. Select "New Inclusion Criteria" and fill in the fields about the criteria. Then select +Add criteria to group and select "Add Condition Occurrence" (or suitable criteria based on your phenotype of choice)

The screenshot shows the 'Inclusion Criteria' section. It displays a list of criteria: '1. migraine cases'. Below the list is a dropdown menu with 'all' selected and a 'having' label. To the right of the dropdown is a 'Copy' button. A sidebar on the right provides options for adding more criteria: 'Add Demographic', 'Add Condition Era', 'Add Condition Occurrence', and 'Add Visit'. At the bottom of the sidebar is a 'Delete' button.

18. On the right select, +Add attribute and select "Add Condition Source Concept" (or suitable criteria based on your phenotype of choice)

The screenshot shows the 'Cohort Exit' configuration screen. At the top, there's a search bar with filters: 'with at least 1 using all occurrences of: a condition occurrence of Any Condition'. Below this, a constraint is defined: 'where event starts between All days Before constraint'. There are two checkboxes: 'restrict to the same visit occurrence' and 'allow events from outside observation period'. A button '+ Add attribute...' is visible. To the right, a dropdown menu lists various filtering options, with 'Add Condition Source Concept' highlighted. Other options include 'Add First Diagnosis', 'Add Age at Occurrence', 'Add Gender', 'Add Condition Start Date', 'Add Condition End Date', 'Add Condition Type', 'Add Visit', 'Add Stop Reason', 'Add Provider Specialty', and 'Add Nested Criteria...'. A red 'Delete Criteria' button is at the top right of the main search area.

19. Then pick with “at least”, “1”, and “using all”.

Under “Condition Source Concept”, select arrowhead for “Import Concept Set” and select the concept set you defined, here `Sandbox_Oct_migraine`.

To note for E, D, and M from [here](#):

Exclude: Selecting this checkbox will prevent that concept from being used in the concept set.

Descendants: Selecting this check box will use the vocabulary relationships to automatically select all descendants. If this option is used in conjunction with the exclude option, it will exclude the current concept and all descendants.

Mapped: Selecting this check box will use the vocabulary relationships to automatically select all concepts mapped to the selected concept.

In the case below, I picked **D** and **M**

Inclusion Criteria

New inclusion criteria

1. migraine cases

migraine cases

enter an inclusion rule description

having all of the following

with at least 1 occurrence of

a condition occurrence of

Condition Source Concept is Sandbox_Oct_Migraine

Import Concept Set

Clear Concept Set

Migraine

Name	E	D	M
Migraine	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

+ Add criteria to group... Delete Criteria

+ Add attribute... Delete

where event starts between All days Before and All days After index start date [add additional constraint](#)

restrict to the same visit occurrence

allow events from outside observation period

Limit qualifying events to: earliest event per person.

To note: to get the image above to show, hover over `Sandbox_Oct_Migraine`

After selection, your window should look like this:

with at least 1 using all occurrences of:

a condition occurrence of Any Condition

Condition Source Concept is Sandbox_Oct_Migraine

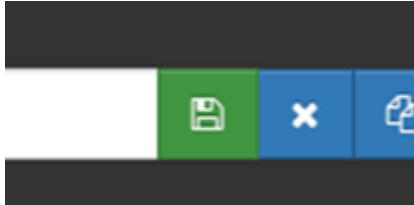
where event starts between All days Before and All days After index start date [add additional constraint](#)

restrict to the same visit occurrence

allow events from outside observation period

+ Add attribute... Delete

20. Scroll up and save your cohort using the “save” button



21. Finally, go to the tab that says “*Generation*” to fill your cohort. This will generally take a couple minutes. It is not dependent on the size of the sandbox.

Tip: Before you start the generation step, I’d highly recommend looking at what is encompassing your concept set. So, pick out the *Concept Sets* tab and explore.

Cohort #304

TestPhenotypeCases

Definition Concept Sets Generation (circled) Reporting Export Messages (3)

Available CDM Sources

Source Name	Generation Status	People	Records	Generated	Generation Duration
▶ Generate (circled)	n/a	n/a	n/a	n/a	n/a

Cohort #958

Sandbox_Oct_migraine_cases

Definition Concept Sets Generation (circled) Reporting Export Messages

Available CDM Sources

Source Name	Generation Status	People	Records	Generated	Generation Duration
▶ Generate (circled)	RUNNING	10/06/2021 2:54 AM	...
Cancel	FinnGen CDM R7				

Cohort #958

Sandbox_Oct_migraine_cases

Definition Concept Sets Generation Reporting Export Messages (3)

Available CDM Sources

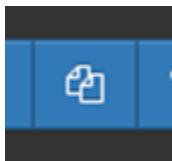
Source Name	Generation Status	People	Records	Generated	Generation Duration
-------------	-------------------	--------	---------	-----------	---------------------

ACTION ITEM: What are the number of cases that you have?

Now, let's get the controls.

For control cohort definition

22. First, let's clone the set that we defined for the case cohort by clicking this button below:



23. You would now have a refreshed screen that looks a bit like this

The screenshot shows the OHDSI interface for defining cohorts. On the left, a sidebar lists various features: Home, Data Sources, Search, Concept Sets, Cohort Definitions (which is selected), Characterizations, Cohort Pathways, and Incidence Rates. Below this, it says "Apache 2.0 open source software provided by OHDSI". The main area is titled "Cohort #959" and shows a message "COPY OF: Sandbox_Oct_migraine_cases". It has tabs for Definition, Concept Sets, Generation, Reporting, Export, and Messages (with 3 notifications). The "Definition" tab is active. A text input field says "enter a cohort definition description here". Below it, a section titled "Cohort Entry Events" contains a dropdown menu "Events having any of the following criteria:" with "Any Visit" selected. Buttons for "+ Add Initial Event", "+ Add attribute...", and "Delete Criteria" are visible.

Rename your cohort now accordingly. I chose to replace cases with controls.

This screenshot shows the same Cohort Definitions page as above, but the cohort name has been changed to "Sandbox_Oct_migraine_controls". The rest of the interface remains the same, including the sidebar and the entry events configuration.

24. Rename your “*Inclusion Criteria*”

This screenshot shows the "Inclusion Criteria" page. A new rule is being defined with the name "migraine controls". The rule description is "1. migraine controls". The inclusion rule description is "migraine controls". Under "having all of the following criteria:", there is one condition: "Condition Source Concept is Sandbox_Oct_Migraine". The "Delete Criteria" button is visible. At the bottom, there is a note about limiting qualifying events to the earliest event per person.

And select “*exactly 0*” of the same Concept set.

25. Click “Save” and go to the “*Generation*” tab to generate your control cohort.

This screenshot shows the Cohort Definitions page again, but the "Generation" tab is now selected. The cohort name is "Sandbox_Oct_migraine_controls". The "Available CDM Sources" table has one row with "Source Name" and "Generation Status" columns. The "Generated" and "Generation Duration" columns are empty.

ACTION ITEM: What are the number of controls you have?

Activity 3.2: Running your custom GWAS

Now that the cohorts are created, we can run a custom GWAS on these. Adapted from [here](#)



1. Click on Applications → FinnGen → Custom GWAS
2. The Custom GWAS tool will open in a Mozilla Firefox browser within the Sandbox.

A screenshot of a Mozilla Firefox browser window titled "FinnGen Custom GWAS - Mozilla Firefox". The address bar shows the URL "https://gwas.app.finngen.fi". The main content area displays the "CUSTOM GWAS" interface. It includes a "DESCRIPTION" section with text about GWAS, a "TOOLS" section with a paragraph about R and I, and a "VISUALIZATION" section featuring a bar chart. Navigation buttons "Help" and "Next" are at the bottom.

- Select your case and control cohort according to how you've named them in the previous activity. You can also type the name out in the dropdown box in order to do a quick search like below.

2. Select cohorts

If you can't find suitable cohort, create a new one in Atlas.

Case cohort

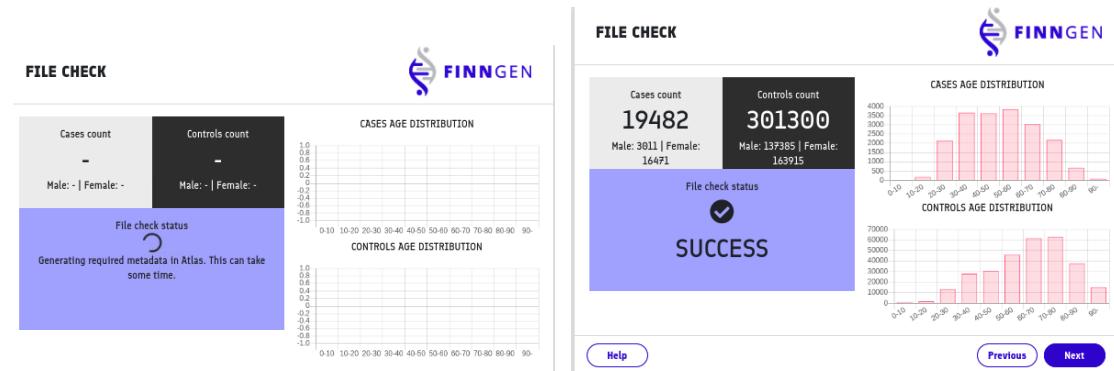
- CRSwNP and asthma and inttolerance ? Argyro B_V
- Sandbox_Oct_migraine_cases
- Sandbox_Oct_migraine_controls

Once you are done with your selections, you can click on next.

- Once you have clicked next, the Custom GWAS tool will do a *File Check* that will provide you with summary results e.g., counts, age of cases and controls. This step may take awhile.

While this is running, perhaps talk about why would it be important to adjust your model with the covariates that are used which are Age, Sex, 10 PCs, and genotyping batch.

ACTION ITEM: If genotyping batches are important, what are the distinguishing bins for the batches in terms of cohorts? Clue can be found [here](#)



- Once you click next, fill in the descriptions

ANALYSIS NAME AND



DESCRIPTION

Give your analysis and description. Your results will be published in PheWeb and Library Green so clear analysis name and description will help other users to interpret your analysis.

Phenotype name

Short tag name to be used as the folder name for results (no spaces or punctuation), for example "T2D"

SandboxOct

Analysis description

Short description of your analysis, for example "Type 2 diabetes after gestational diabetes".

Sandbox October Migraine Kumar

Cases

Example: "Women who have gestational diabetes and later receive a diagnosis of type 2 diabetes"

People with migraine

- Finally, you can now submit your analysis which should be ready in a few hours

NOTIFICATIONS



Almost ready! After you hit submit, your analysis will be published in PheWeb and Library Green tomorrow. You will also be notified when your analysis is scheduled and when it is completed.

Send me a notification when analysis is ready

Notification address:

veerapen@broadinstitute.org



Processing time is 24 hours

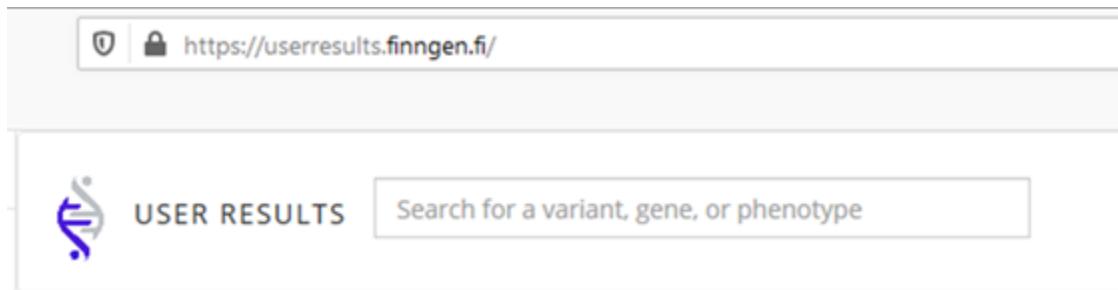
Help

Previous

Submit

- Your results can be viewed outside the Sandbox in the [FinnGen User Results](#) page.

- Sign in with your FinnGen user account



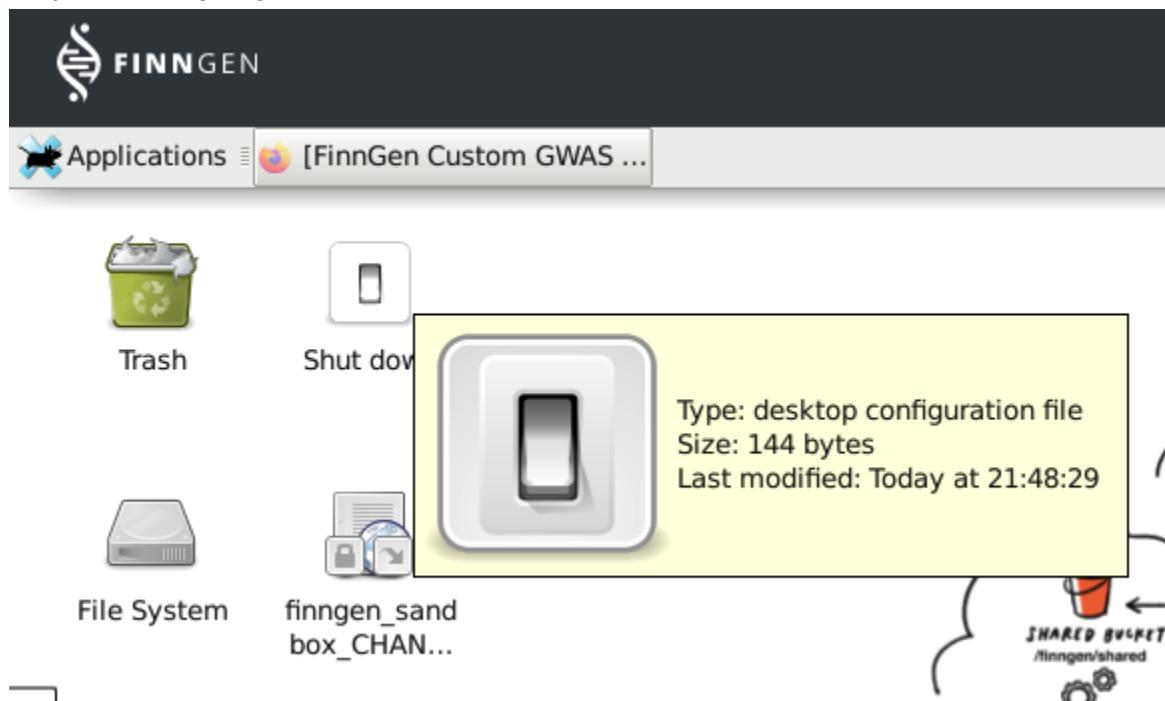
- Search for your phenotype name

Voila!

To note, for ACTIVITY 4, everything you save on your IVM will be there when you return regardless of whether you shut down or sign out.

Activity 4: Shutting down your IVM

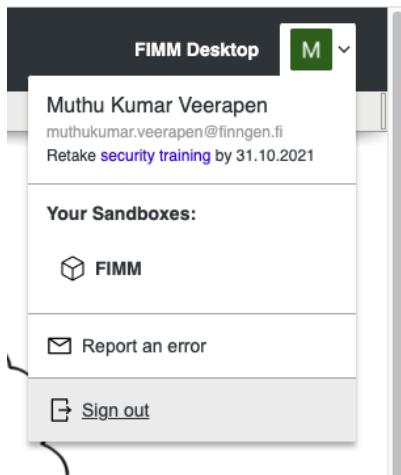
On your landing page, double click on the switch icon



OR

Activity 4: Keeping your IVM running

You can disconnect from your session by either logging out of the Sandbox



Click your name in the top right corner of your window, and select “*Sign out*”

OR

Closing your window

If you want to re-access this session by signing out or closing window, you would just have to repeat Activity 1 but this time, you will immediately be brought to the machine of your choice.

These IVMs will continue running for 48 hours before being forced to shut down. Always a good practice to shut down instead of leaving it running.

Additional Resources

[FinnGen Members Area](#)

[FinnGen Analyst Handbook](#)

[Sandbox Tutorial Site](#)