Submitting stuff to the SRA

You should follow the instructions on their website. Below is a few tips.

- Keep the files on the HPC, you don’t need to download them to your laptop.
- You can fill out the spreadsheets on their web tool, or you can download and re-upload. Your choice. There’s at least two of them.
- Yes, I think they like de-multiplexed fastq, not just the off-the-machine fastq.
- For one of these, you’re going to need to specify sample IDs for different files. Thing is, they also need to be different in places other than the sample IDs. I did this by specifying two columns of ‘timepoint’ and ‘treatment’ for my samples.
- You’re going to need to insert the filenames. Get onto the HPC, find where they are, and you can do something like

  ```bash
  ls /scratch/cgsb/gencore/out/Gresham/2017-01-06_HGGNWBGX2/new/*_[wq]* | cut -d / -f 9 > list_of_files.txt
  ```

  to get a text file with their names. Above, that's everything with _w or _q in the name, because I don't want the undetermined fastq. `sftp` that to your local, and you can copy and paste it into the form.

- Once you've got that done, you need to upload the actual files. This is how to FTP it from the HPC:
  - look for that center FTP thing, and click on that link
  - it'll take you back to the same page, again click on the FTP link, and you'll see that it made you a short-time password
  - then it has some pretty explicit instructions, including a new temporary username and password, and specific instructions about making a certain folder and putting your files in there. Wait on this.
  - Log on to the HPC, you check out where your files are, and you can gzip them or not. You might not want to, it's going to take forever and they need to unzip them anyways. Make sure it's the same filename as you specified above.
  - Start a tmux session by typing tmux. It should look very similar, but with a green bar. This is important for letting the session run while it's uploading, without you having to have it open on your box. (more details here https://linode.com/docs/networking/ssh/persistent-terminal-sessions-with-tmux/ or ask Darach or Pieter)
  - Go ahead and type `control+b` then `%` to open another terminal panel, you can switch between them with `control+b` then `n`.
  - In one of these terminals, do the FTP command it says on the sra submission website as linked above (https://submit.ncbi.nlm.nih.gov/subs/sra/). No sftp, it's ftp.
  - Once you're logged in (copy and paste the username and passwords given), then change directories as specified.
  - Make a folder, name doesn't matter, and change into it.
  - Then, enter `prompt`. This turns off the interactive mode for the next step.
  - Then, enter `mput *` to upload everything in your current directory. If it's more complex, do something like `mput "gz" or `mput /scratch/cgsb/gencore/out/Gresham/2017-01-06_HGGNWBGX2/new/*_[wq]"` or whatever. See, it's like a "multiple put" instead of "put"ing one at a time.
  - It should just upload all of them. This will take a while, so you can detach the tmux to close the session (default is `control+b` http://github.com/tmux/tmux/wiki/FAQ), and then immediately the key `d`, and that should detach the session. Alternatively, you can switch to the other window `control+b` then `n`, and just enter `tmux detach` to detach the session.
  - It'll take you back to the default hpc, then you can log off. When you want to log back on, log on and enter `tmux attach` to get the session back on check it. It'll take a long time.
  - You can also just leave it open for however long it takes.

- Once it's uploaded to their FTP, then you can exit everything. Then go back to the submission on the SRA website, continue it, and in Step 7 select the folder you uploaded to. It'll take a while for them to process the samples.

You might get an error. Let's say that they folks who do the bioinformatics didn’t inspect their new demultiplexer outputs enough before running it on your samples, and it stuck newlines everywhere. You'd have to remove those, or SRA is going to choke on the format error. Watch out, and be nice to the tech support at NIH.