Hack a thon technical specs

1. Create virtual environment

* First create directory where virtual environments will live
  + Mkdir environments
* Change directory into environments, create virtual environment, install packages
  + virtuaenv project1\_env (project1\_env is the name of the VE)
  + source project1\_env/bin/activate
  + (project1\_env)astylianou900045@hpcc-1:/mnt/hnas/bioinfo/users/Alexandrea/hack-a-thon/enviorments$ which python
    - /mnt/hnas/bioinfo/users/Alexandrea/hack-a-thon/enviorments/project1\_env/bin/python
  + (project1\_env)astylianou900045@hpcc-1:/mnt/hnas/bioinfo/users/Alexandrea/hack-a-thon/enviorments$ pip list
    - argparse (1.2.1)
    - pip (1.5.6)
    - setuptools (5.5.1)
    - wsgiref (0.1.2)
  + pip install pyvcf
  + pip list
    - argparse (1.2.1)
    - pip (1.5.6)
    - PyVCF (0.6.8)
    - setuptools (5.5.1)
    - wsgiref (0.1.2)
  + deactivate (now that we are done creating the virtual environment, deactivate and start coding)

2. Activate VE and run program

* cd into where your projects are
  + :/mnt/hnas/bioinfo/users/Alexandrea/hack-a-thon
* Create a shell script to activate VE and run python program

#!/bin/bash -x

set -o errexit

source enviorments/project1\_env/bin/activate

python Part1\_RIT.py

3. Create VCF file

* Download .bgz file and unzip it
  + gunzip -c gnomad.exomes.r2.1.1.sites.21.vcf.bgz > gnomad.exomes.r2.1.1.sites.21.vcf
* run bash script to convert .vcf into fastA file
  + \mnt\hnas\bioinfo\users\Alexandrea\hack-a-thon\fastAchr21.txt

4. blast

* Made blast db:
  + makeblastdb -in chr21.fa -dbtype nucl -parse\_seqids -out chr21dbW
  + which created these files:
    - chr21.fa.nsq
    - chr21.fa.nhr
    - chr21.fa.nin
* remove simple sequences
  + blastn -query fastAchr21.fa -out fastAchr21.fa.txt -db chr21db -dust yes <-this can be disregarded, appears that dust is default
* blast chr21
  + blastn -task megablast -query fastAchr21.fa -db chr21db -best\_hit\_score\_edge 0.1 -gapopen 5 -gapextend 2 -reward 1 -penalty -1 -word\_size 11 -out blastchr21\_output.txt -outfmt "6 qseqid sseqid qcovs stitle qlen qstart send qseq sseq evalue bitscore score length pident nident mismatch positive gapopen gaps ppos sstrand"