Appsoma-based pipeline development
Vidhoon and Naveen -
Precursors for doing step 0:
1) Create an account at appsoma.com by logging into the UT system with your EID and make an account (use your EID as your appsoma user name)
2) read the "help" tab information - especially the last bit on "How Appsoma Works" and all the other sections EXCEPT the Lanugage API which you can read as you need it.
3) Access some code from other people: "Account" -> "Add new collaborator" - add jjc8993 and drdna and appsoma and bennigoetz.
Test projects:
0.a) setup, test, and improve the Blast app: https://appsoma.com/code/jjc8993_blastn.bash - specifically, make this run "tblastx" as well as "blastn"
0.b) setup, test, and improve the TopHat/Tuxedo suite and STAR run on an arbitrary dataset: https://appsoma.com/code/jjc8993_rnaseq_mult_h ead.py - specifically, make the two of these programs run concurrently on a "real" dataset.
2 "real" projects:
1) stand-up projects in engineering, physics, CS, stats, business, math. Start with stats (appsoma)
2) RNA-seq analysis in human - see 0.b. (DPAC)
These will share the following significant research challenges:
1) What is the best data model to use in Appsoma? This will encompass simple questions such as where and how to store metadata fields as well as much more complicated questions such as intelligent data replication algorithms, automatic metadata generation.
2) Community-based analytical kernels - a system which takes a panoply of analytical methods, determines which might possibly be applicable to the data given, runs them all in parallel, presents the results and ideally cross-compares the results and makes suggestions of what to do next. A positive example of this already working is MS Blender: https://appsoma.com/apps/demo_msblender.bash which is based on work from this paper: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3128686/?report=reader
Please add your questions/thoughts/ideas/approaches below or as child pages!
Best Practices on Appsoma
Exome pipeline