17 August 2016:

Branch v1_errorcatch

First: commit

- branch off of v1 to start working on step 1 exception catching

Branch v1

Second: commit

- got rid of excel writing (more memory efficient since we are already saving csv; less data lost in case of error/break; simpler)
- shortened time of sql vars
- modified run / streamline and adjusted to modified functions and classes
- annotations to Counts and run
- ran and debugged run (works for selected indexes of nate's data) !!
- (just need to adjust filtered file thing - not updating...)

First: commit **ae1ea2ed921c84b5552686a151d2d45ddd2962ad**

Checkpoint for Counts edits with the following changes

- separate qgm_counts_df construction and filtered_df construction for modularity
- "construct_qgm_counts_df" is the same as the first half of defunct 'construct_df'
- new "construct_filtered_df" --- takes qgm_counts_df, counts no. molecs and reads PF, named 'molecs' and 'reads' henceforth, and --- filter data by molec(PF) > 0 and all features non-null --- plus some formatting for user

16 August 2016:

Branch v1

Second: commit **bb5de4525e3752a47528d651e7963a25a34b08fb**

Checkpoint for finishing Counts (ugh finally) and fixing Index

- in Index.count_reads, should be type(counts[key])==list (not np.array -- don't know why it doesn't work. It's probably fixable but i can't deal with it right now because I need to finish all of this other schtuff)
- in Counts, constructed generators to manipulate df as static methods: --- 'parse_qgm_key' from input tuple --- 'calculate_count_minscore' to count total and pf reads, bool(molec_passed) according to minscore --- free memory: drop qscore seqs, and eg return intermediate qgm_df to outer function (instead of as property) --- count no. molecs and reads which PF per qg --- set Counts.qg_df property
- added xlsx, csv saving functionalities
- in progress: amending 'run' to align with other changes, and streamline/declutter; annotations

(Next: catching exceptions!!!!!!) also, maybe format motif function to ease user input requirements (e.g. instead of {ATCG}, convert "." in function to clean up user input)

First: commit **98b5419dad3090c20330b43bb25575f77fe1121e**

Checkpoint for finished Index, Counts in progress

- finished writing Index and surface debugging / catching
- cleaned Counts fn construct_df (with helpers parse_qgm_key and calculate_count_minscore)
- in progress: debugging construct_df, re-writing counting qg molecs & reads

15 August 2016:

Branch v2

First: commit **8d64e20ace3c1344c216b29a2eae8d71469afac3**

- set Index to call open_gz, cleaned Index obj:
- Added read_ref option for UX flexibility
- Streamlined count_reads logic and flow, calling motif_search
- Major modifications to motif search; generalize search and feature extraction for each feature and to minimize downstream conditional statements