Follow up:

1. discussed the outline and OK with it
2. we sent the enhancer list, and waiting for the comparison
3. how to handle the variant transfer – JZ to ask Kevin
4. follow up with Michael to ask for k mer (ANS)
5. MCF7 or MCF10 – ANS to Brad Bernstein
6. What is the complement for breast? – can we use REMC paper as a guide?
7. LnCap cell-line – histone marks may not be in ENCODE
8. Shirley’s list of things that they have for ENCODE cell-lines – JZ
9. Doodle poll for time for ENCODE cancer call – Kevin, Robert Klein, Shirley Liu
10. JZ – resending bits of outline back to Kevin and summarize most of the points in an email

heather jason michael

ANS, JZ

encode cancer paper, enhancer-seq

q&a

1) What is the difference between starseq and enhancer-seq?

enhancerSeq is typically not performed on a whole genome basis. Typically, they capture elements using another assay (like open chromatin regions) and then perform STARR-Seq with these regions. It is more similar to CapSTARR-Seq in some ways.

2) Enhancer-Seq: what promoter will be used in their assay ?

They used FCP1 (I think) promoter. They will continue to use the same promoter. They think MCF7 does not work too well with whole genome and they would prefer 10K elements that can be tested due to experimental difficulties with transfection to the MCF7 cell-line.

3) previous results for this promoter because promoter can influence whether an enhancer is positive in the assay (shows assay sensitivity - and avoiding false negatives). Can we get some training data for this promoter in similar cell lines.

They would not give us the results from previous experiments but would rather check whether our predictions make sense by comparing to previous experimental results. We will send our predictions to them (already done).

4) What is the length of the region being tested in enhancerSeq?

500 bp

5) Of 10k enhancers, 100k regions, what percentage of active ones?

It could be up to 6k, open regions (later)

6) number of samples for variant analysis?

85 samples TCGA, er- sample, 75 of the sample, 44~45 er-, 30 er+

7) how many variants gone test?

similar fashion with fuseq, 250 variants

8) center of the variants:
They would prefer if the variant is centered on the enhancer - they will take 250 bp on each side of the region in their assay.

9) Timeline

enhancer-seq: order the library, <7 weeks get our hands on, 4 weeks more generate the input library (may be a few weeks longer due to sequencing), 2~5 weeks have the updates on that

4 + (7+4) + 2 + 6 = 23

how long: a couple of weeks 6 weeks to get the synthesize, once sytem

10) Do we want to use promoters on the enhancerSeq assay?
Prefer using CRISPR knockdowns for this rather than enhancerSeq assay. So in the 10K regions, no promoters will be tested.

strategy:

MG: preliminary discussion with ppl

KW: TF knock out, or cripser the out, do expression profiling to see the predicted changes occured.

MG: how many

KW: between 5 and 20, RNA-seq probably. More enhancers-seq > star-seq, more than 20k possibly

M… cell line (more normal than MCF7)

**Suggestion (after call): HMEC is a breast cell line from ENCODE (not cancer). I am already making enhancer predictions for this cell line to compare to the enhancer predictions from MCF7. Can we use this cell-line too for the study - to compare normal breast cells?**

**050-0011182-2552**

**Carolyn: all TCGA data will be realinzed in 6~9 month**

**Robert: GWAS paper plans**

**MG: also colin and MK**

**RC: enhancers how to be useful to GWAS and what is the validation assay.**

**MG: mcf7 and**

**CML K562**

**MCF10 also related to breast cancer (also some encode data, more attractive than mcf7, more normal than mcf7)**

**Lung cancer cell line,**

**other opportunity for ppl who has other enhancer predictions.**

**RC: GWAS & cancer? no specific, germline variants, how common**

**annotations of recurrent mutations in breast cancer**

**Carolyn: around 900 for WGS**