**Methylation Biomarker Project for Breast Cancer**

We will follow the pipeline what we have established in our previous ESCC project. In breast cancer project, I will apply bioinformatics and biostatistics method to select 50 most potential differential methylation regions (DMR) in breast cancer (BRCA). These 50 regions will have at least 5-7 tumor suppressor genes so that we can prepare the second manuscript. Meanwhile, we will select 10 novel DMR so that we can claim we identified new breast methylation biomarker. Finally, we will select 10 DMRs who have potential important functions, such as transcriptional factor binding site (TFBS) or miRNA binding site (miRNA-BS) so that we can conduct some research like the third manuscript. According to our previous research, 50% of 50 regions will be significantly differential methylated in our project while other 50% might be failed. The reason why these 50% regions will be failed include:

1) TCGA dataset is white people, we repeat these biomarkers in Chinese population

2) Subtype difference. BRCA have many subtypes which maybe different between our samples and TCGA

3) Sampling bias, we only selected 150 pair samples which might have certain bias.

4) Genetic background difference, as GWAS shown genetic etiology Chinese and White have difference.

What we will do:

1. Evaluate the sensitivity, specificity and AUC for each DMR/biomarker
2. Identify a best prediction model to distinguish cancer and adjacent normal
3. Identify outcome associated biomarker
4. Identify TNM related DMR
5. Identify other clinical information related DMR

Sample collection requirement:

1. Age, gender, BMI, subtype, smoking, drinking, family history, TNM and overall survival time (OS)
2. TP53, PIK3CA and GATA3 mutation: Yes or No
3. Samples with other omics data should give high priority to enter the project (WES or RNAseq)
4. Other clinical information such as WBC?, AFP?, other routine checking for BRCA
5. Staining will be another interesting information. Tumor purity can be evaluated by staining
6. Max Tumor size? Tumor number?
7. DNA: 100 paired BRCA-adjacent, 5 BRCA cell line (before and after 5-AZA treatment for 3 days)
8. cfDNA: 40 BRCA cfDNA and 40 normal cfDNA (age, BMI, drinking, smoking matched samples)
9. totally: 300 DNAs (100 BRCA, 100 BRCA-adjacent, 5 cell line before, 5 cell line after, 45 paired cfDNA)
10. DNA, RNA, protein should be collected simultaneously after 3 day treatment of 5-AZA since mRNA and protein might be used for RT-PCR and Western blot

[Targeted bisulfite sequencing identified a panel of DNA methylation-based biomarkers for esophageal squamous cell carcinoma (ESCC)](javascript:void(0))

W Pu, C Wang, S Chen, D Zhao, Y Zhou, Y Ma, Y Wang, C Li, Z Huang, ...

Clinical epigenetics 9 (1), 129

[Identification of hyper-methylated tumor suppressor genes-based diagnostic panel for esophageal squamous cell carcinoma (ESCC) in a Chinese Han population](javascript:void(0))

C Wang, W Pu, D Zhao, Y Zhou, T Lu, S Chen, Z He, X Feng, Y Wang, ...

Frontiers in genetics 9

[Epigenetic silencing of ZNF132 mediated by methylation-sensitive Sp1 binding promotes cancer progression in esophageal squamous cell carcinoma](javascript:void(0))

D Jiang, Z He, C Wang, Y Zhou, F Li, W Pu, X Zhang, X Feng, M Zhang, ...

Cell Death & Disease 10 (1), 1

大体的时间表如下：

1. 和天昊签署好合同后，我就可以把DMR区域发给天昊，他们负责设计并评估引物。
2. 我一般发给他们100个区域，他们一般只能设计成功50个，整个引物评估大约需要经过多轮的协商，因为很多区域可能无法设计引物进行检测。往返过程大约需要2-3周才能把最终的50个区域确定。剩余50个区域其实也很有价值，我们可以在自己实验室进行攻关，天昊因为是做multiplex PCR所以对区域之间组合的要求比较高，如果我们自己lab单独做，则没有这些严苛的要求，可以进行逐个评估。
3. DMR区间和引物评估完成后，我们就可以送DNA到天昊，他们开始进行甲基化DNA的亚硫酸盐转化，及DNA甲基化检测，以及出具报告，这个过程大约需要4-6周时间。
4. 收到他们的甲基化检测结果后，我可以在2周之内提供两篇文章（ESCC中的第一篇和第二篇）所需要的所有图表。然后准备文章的撰写，专利的撰写及投稿
5. 自第三步开始，我们就可以准备文章初稿，背景部分，method部分，8周时间准备这两部分绰绰有余。注意就是把breast cancer中涉及甲基化的文章梳理一下，把高分的文章拿出来总结一下放到background中以及留着在discussion中进行比较讨论。空着method和讨论部分，等结果到位，我提供所有图，表后即可快速成文。
6. 现在是4月，如果一切顺利，估计最快7-8月份就可以投稿。