AD_____

Award Number: W81XWH-11-1-0582

TITLE: Progression of structural change in the breast cancer genome

PRINCIPAL INVESTIGATOR: Hartmaier, Ryan J (Postdoctoral Associate)

CONTRACTING ORGANIZATION: University of Pittsburgh
Pittsburgh,PA..152134-3320

REPORT DATE: August-2013

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
August-2013	Annual Summary	01 August 2012 – 31 July 2013

4. TITLE AND SUBTITLE
Progression of structural change in the breast cancer genome

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-11-1-0582

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Hartmaier, Ryan J

E-Mail: hartmaierr@upmc.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Pittsburgh
3520 Fifth Ave
Pittsburgh PA 15213-3320

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

As our basic understanding of the human genome evolves, we are beginning to appreciate that it is not a static entity but rather a plastic one acquiring *de novo* mutations and structural changes. A number of recent studies suggest that breast cancer is initiated through disrupted DNA repair processes, leading to a destabilized genome, in turn promoting a heterogeneous primary lesion from which a/many subpopulation(s) acquire general or organ specific metastatic potential. I aim to identify and characterize the specific mutations that at acquired during breast cancer metastasis. To do this paired primary and metastatic breast cancer samples have been obtained and used for targeted and genome-wide analyses. Large insert mate-pair sequencing has been completed for 7 samples from 2 patients. Analysis is ongoing but many primary-metastasis shared and metastasis specific structural changes have been identified. Additionally, I present strong evidence that NCOR1 and a number of other genes are mutated specifically in ER+ disease and thus represent ideal targets for endocrine resistance research. Attached herein, I provide a detailed progress report for this project.

15. SUBJECT TERMS
breast cancer structural change, copy number variant

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	15	19b. TELEPHONE NUMBER *(include area code)*

Award: W81XWH-11-1-0582

PI: Hartmaier RJ

Report: Annual Report (1 Aug 2012 – 31 July 2013)

Table of Contents

Introduction

Genomic instability is an "enabling characteristic" of cancer allowing for the acquisition of mutant, cancer-promoting phenotypes (i.e. sustained growth signaling, activation of invasion/metastasis). Large-scale cancer sequencing studies, such as The Cancer Genome Atlas (TCGA), provide an excellent resource to identify genetic events that are driving cancer. However, many passenger, non-driving events are also identified. To help interpret these results, huge sample numbers and/or complicated pathway prediction models are required. Complicating this analysis further, we are beginning to appreciate the extensive genetic heterogeneity within a tumor, like much a result from independent passenger mutations occurring throughout the evolution of the primary tumor. The approach I have proposed in this fellowship is to leverage paired tumor samples from the same patient to uncover events that have been sustained or uniquely acquired during metastatic progression. Here I describe my progress in obtaining the appropriate tissues, characterizing a candidate copy-number variation, and conducting genome-wide rearrangement detection.

Body

Task 1: Gain necessary approvals and receive tissue samples needed for the study (months 1-6)

> *1a: Panel paraffin embedded blocks of breast cancer progression (whole blood, DCIS, ductal carcinoma, metastatic lesion) (20 samples from each tumor for pilot study which will determine numbers needed for full study)*

Necessary tissues have been obtained. There is nothing more to report.

> *1b: Matched blood, primary breast cancer, and metastatic lesions (3 tissues from 10 individuals)*

We continue to collect paired normal, primary, and metastatic breast cancer samples. The bulk of the paired, frozen samples have been collected although the rapid autopsy program continually accrues patients and should more samples become available, they will be added to the studies. Many more FFPE pairs have been collected. These samples are not appropriate for mate-pair sequencing (high molecular weight DNA is required), however, they are available for targeted studies (validation) and are appropriate for other sequencing technologies that we are pursuing outside the scope of this fellowship (e.g. whole exome sequencing). See Table 1 for an updated list of matched samples and the current state of analysis.

Table 1: Summary of paired, frozen breast cancer tissues and current status of analysis. Analyses performed include: Affymetrix Genome Wide SNP Array 6.0 (Affy6.0; genome-wide copy number), Ion Torrent Ampliseq 2.0 beta (Ampliseq2.0; targeted mutations), NanoString Copy Number Variation beta (NanoString CNV; targeted copy number), bisulfite converted RainDance Technologies targeted amplification (RainDance; targeted methylation), and large insert mate-pair sequencing (Mate-Pair; genome-wide rearrangements, copy number, and mutation). IP=in process, To send=will begin in ~1 month.

Patient	Sample Type	Site	Tumor Type	Tissue Type	Affy6.0	AmpliSeq2.0	NanoString CNV	RainDance	WES	3-5, 5-8, 8-12kb Mate-Pair	40kb Mate-Pair
RJH-MET-1	Tumor	Breast	Primary	Frozen					To do	To do	
	Tumor	Breast to Lymph Node	Metastatic	Frozen					To do	To do	
RJH-MET-2	Normal	Breast		Frozen					To do	C	
	Tumor	Breast	Primary	Frozen					To do	C	
	Tumor	Breast to Lymph Node	Metastatic	Frozen					To do	C	
RJH-MET-3	Normal	Buffy Coat		Frozen		C	C				
	Normal	Spleen		Frozen	C					C	
	Normal	Lt Breast		Frozen				C	To do		
	Normal	Rt Breast		Frozen							
	Tumor	Rt Breast	Primary	Frozen	C	C	C	C	To do	C	
	Tumor	Lt Breast	Local Recurrence	Frozen	C	C		C			
	Tumor	Lt Breast	Local Recurrence	Frozen					To do	C	
	Tumor	Breast to Liver	Metastatic	Frozen	C	C	C	C	To do	C	To do
	Tumor	Breast to Thoracic bone	Metastatic	Frozen	C	C	C				
	Tumor	Lung	Metastatic	FFPE							
	Tumor	Lung	Metastatic	FFPE							
	Tumor	Lung	Metastatic	FFPE							
RJH-MET-4	Normal	Rt Occipital		Frozen	C	C					
	Normal	Rt Occipital		Frozen							
	Normal	Lt Breast		Frozen				C	To do	To do	
	Normal	RLL Lung		Frozen							
	Normal	Lt Breast		Frozen							
	Normal	Liver		Frozen							
	Tumor	Lung	Primary	FFPE					To do		
	Tumor	Lung	Primary	FFPE							
	Tumor	Lt Breast	Local Recurrence	Frozen	C	C		C	To do	To do	
	Tumor	Lt Breast	Local Recurrence	Frozen				C			
	Tumor	Lt Breast	Local Recurrence	Frozen							
	Tumor	Breast to Lymph Node	Metastatic	Frozen	C	C			To do	To do	
	Tumor	Breast to Liver	Metastatic	Frozen	C	C		C	To do	To do	To do
	Tumor	Breast to Rt Occipital	Metastatic	Frozen	C	C		C	To do	To do	
	Tumor	Lung	Metastatic	FFPE					To do		
	Tumor	Lung	Metastatic	FFPE							
	Tumor	Lung	Metastatic	FFPE							
RJH-MET-5	Normal	Breast		Frozen					To do	To do	
	Tumor	Breast	Primary	Frozen					To do	To do	
	Tumor	Breast	Local Recurrence	Frozen					To do	To do	
RJH-MET-6	Tumor	Breast	Primary	Frozen					To do	To do	
	Tumor	Breast	Local Recurrence	Frozen					To do	To do	
Completed					10	10	4	9	0	7	0
Total	39				10	10	4	9	21	19	2

Task 2: Determine impact of NCOR2/SMRT CNV on breast cancer progression (months 1-24)

2b: Better determine the region the CNV encompasses in lymphoblastoid cell lines and breast tumors previously identified to harbor CNV (months 1-3, samples have already been approved for use)

Previously published data and my own preliminary data suggested that a germline CNV exists in the NCOR2 locus. This includes a number of SNP array studies, fosmid mate-pair end-sequence profiling (ESP), and QPCR based copy number analysis.

To better determine the extent of the copy number variation and to identify additional samples with the change, I previously obtain DNA from 8 individuals identified in previously published reports to harbor a copy number change in NCOR2 and a panel of 24 additional individuals. These samples were[1] tested extensively with two different QPCR based copy number technologies (Life: Probe/RNaseP & Qiagen: SYBR/multi-copy reference) and many assays spanning the entire NCOR2 region. One of these assays, located very close to a fosmid ESP identified deletion, showed evidence of a germline deletion. Analysis with further assays was unable to confirm this finding (Figure 1). Further, although these assays can reliably detect high-level amplifications, it became apparent that the variation in any one assay, combined with the lack of a proven positive control, makes the detection of relatively small changes in copy number (1-2 copies) extremely difficult.

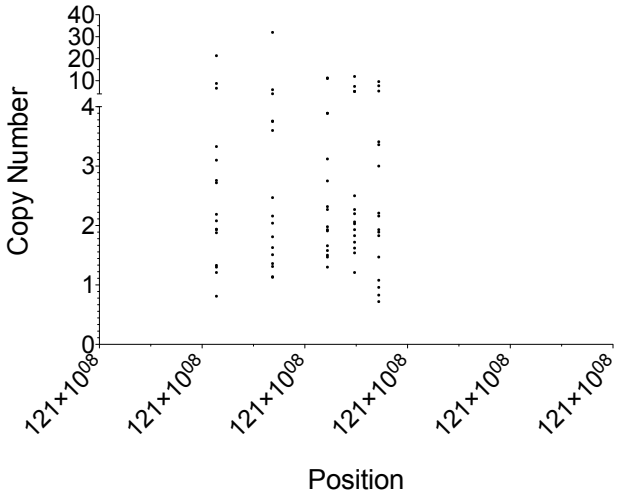

Figure 1: Individual QBiomarker CNV assays. Each dot represents an individual DNA sample for a given assay. Calculated copy number shows a wide rage making it impossible to confidently call a CNV in this region.

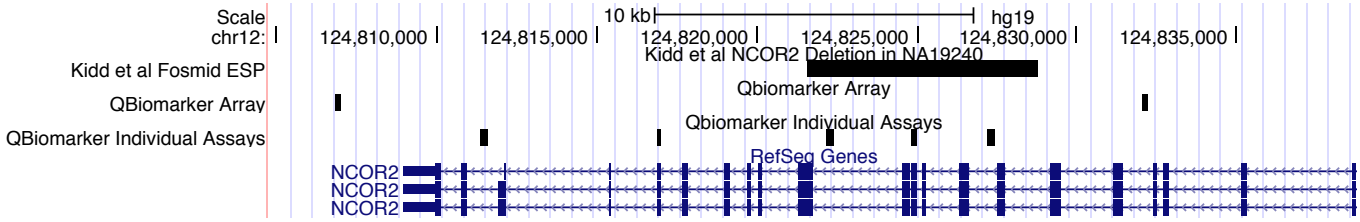

Figure 2: Location of individual and arrayed QBiomarker assays. Five individual assays were designed in the area closest to the array assay previously suspected to be in a deleted region. Three of these assays were specifically chosen to overlap with the deletion identified by Kidd et al.6

Since lymphoblastoid cell lines (LCLs) are available for each of the samples tested, mRNA QPCR was conducted as an alternative approach to determine whether the samples contained a deletion in NCOR2. LCLs corresponding to the three DNA samples showing evidence, as well as an additional 5 samples not showing evidence for a deletion in NCOR2 were obtained. mRNA was isolated from cells during normal culture conditions and QPCR was performed for NCOR2. There was no significant difference in NCOR2 expression based on the putative CNV (Figure 3). Although this does not exclude the possibility of a heterozygous loss with compensatory upregulation of the sister allele, it makes extremely unlikely that a CNV is present in these cell lines that disrupts NCOR2. As a final test, a digital PCR system (BioRad) has recently become temporarily available. This technology is claimed to overcome the inherent variation in QPCR based CNV assays. The same assays and samples previously tested will be run on the digital PCR system in order to have a firm answer to the question of the existence of a germline CNV in NCOR2.

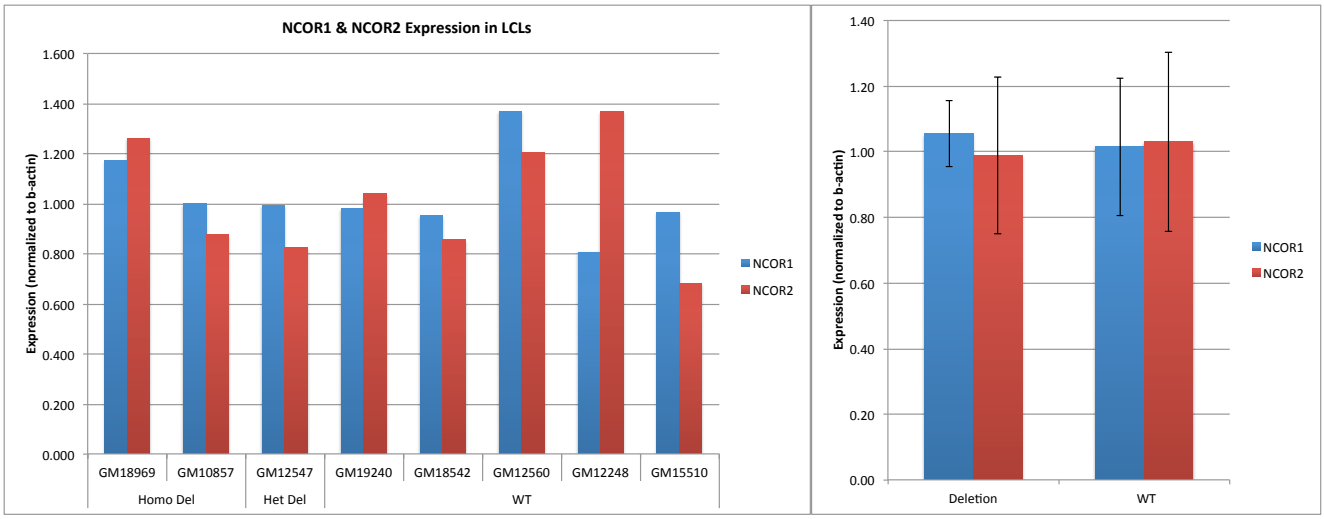

Figure 3: NCOR1 & NCOR2 expression in LCLs suspected to contain a deletion in NCOR2. (left) expression for each individual and (right) averaged based of suspicion of NCOR2 deletion. No significant difference was observed.

A number of large, breast cancer sequencing and high-throughput studies were published in the last year. This includes a report of a novel splice form of NCOR2 that was identified in a tamoxifen resistant variant of ZR75-1 breast cancer cell line and is associated with tamoxifen resistance in patients [2]. The authors do not report on the genomic architecture of NCOR2 or whether a deletion could be responsible for the altered transcription. Thus, NCOR2 remains a great candidate for further study in tamoxifen resistance. The large sequencing studies have found that NCOR2 is mutated in breast cancer at low frequency[3]. Although this provides some evidence that NCOR2 is indeed undergoing some mutational pressure during carcinogenesis, it also shows that if it is present, it is at very low frequencies difficult to confirm or use in a clinical setting.

These sequencing studies have also reveled recurrent mutations and copy number loss of NCOR1, the homologue of NCOR2. It was shown from 100 whole genome sequences that NCOR1 is mutated at ~5-8% frequency. I postulated that genes like NCOR1 and NCOR2 are critical for maintaining tamoxifen antagonism and/or suppressing ER activity. Therefore mutations that occur during tumorigeneis should be selected for specifically in ER+ tumors. To confirm this and identify other potential anti-estrogen resistant mutations, I examined the publically available mutation and copy number data from Stephens et al and from The Cancer Genome Atlas (TCGA). By comparing the mutation rates within a given gene between subgroups, passenger mutations should occur randomly in both groups, whereas biologically relevant mutations should be enriched. Using the TCGA data, when all non-synonymous mutations are considered (missense, frameshift, nonstop, or splice site variation), 26 genes are significantly enriched in either ER+ or ER- disease. Of those, 6 are enriched in ER+ disease. TP53, a gene known to be frequently mutated in ER- disease, serves as a good control. Others like GATA3, PIK3CA, MAP3K1, and CDH1 are all known to be mutated more in ER+ disease. When the missense mutations are excluded (leave only mutations with obvious deleterious effects), this list is further refined to 7 genes. Interestingly, NCOR1 now reaches significance since it frequently undergoes frameshift mutation in ER+ tumors. These obvious deleterious mutations occur at ~3% in ER+ disease but nearly never occur in ER- disease. This effect is strengthened when combined with CNV loss in the region as nearly all of the tumors that harbor a mutation also have a CNV loss – a classic sign of loss of heterozygosity (data not shown). GATA3 is an even more extreme example harboring frameshift or splice site mutations in over 10% of ER+ disease and never in ER- disease. Together these data show that NCOR1 is lost specifically in ER+ disease and that this could represent a potential pathway to hormone therapy resistance – likely in combination with some of the other mutations identified here.

2b: Develop and test FISH probes to detect SMRT CNV (months 4-6)

The previous QPCR based results have not been able to accurately identify if a CNV exists in NCOR2 nor its location. The digital PCR should give a reliable result and may help identify the region involve in the CNV. If a FISH probe can be designed, it will be used as the 'gold-standard' for validating NCOR2 copy number.

2c: Conduct SMRT CNV FISH in pilot breast cancer progression samples (months 7-9)

It is clear from published studies at this time that NCOR2 is not altered at an appreciable frequency in breast cancer. The resources and number of samples required to even detect this change in tumor samples are not feasible and thus this sub-aim will not be continued.

2d: Conduct full SMRT CNV FISH study (months 10-18)

Not started.

2e: Expose CNV-negative lymphoblastoid cells (previously obtained and approved) to ionizing radiation and test for SMRT CNV acquisition (months 19-24)

Not started. Based on the extensive evidence for NCOR1 loss in breast cancer and the many similarities between NCOR1 and NCOR2, this aim will be modified to examine the ability and frequency of the NCOR1 loci to be lost under stressful conditions.

Table 2: Genes significantly enriched for mutations based on ER status of tumor. Data was accessed via TCGA public access. All mutations causing a change in amino acid or translation (missense, frame-shift, nonsense, nonstop, splice-site), that occurred at least 5 times in ER+ patients, and were enriched based on Fisher's exact test ($p<0.05$) are shown.

	ER+			ER-			
Hugo_symbol	Mutated	Not-mutated	%	Mutated	Not-mutated	%	**Fisher**
TP53	119	516	18.74	124	63	66.31	**1.48E-33**
GATA3	74	561	11.65	0	187	0	**1.79E-09**
PIK3CA	209	426	32.91	24	163	12.83	**1.37E-08**
MAP3K1	52	583	8.19	2	185	1.07	**8.51E-05**
CDH1	50	585	7.87	2	185	1.07	**1.37E-04**
ENSG00000198804	8	627	1.26	10	177	5.35	**2.31E-03**
MAP2K4	30	605	4.72	1	186	0.53	**2.99E-03**
AKT1	19	616	2.99	0	187	0	**6.97E-03**
MUC17	14	621	2.2	11	176	5.88	**1.08E-02**
F5	6	629	0.94	7	180	3.74	**1.38E-02**
USH2A	20	615	3.15	14	173	7.49	**1.53E-02**
LRP2	13	622	2.05	10	177	5.35	**1.64E-02**
TTN	78	557	12.28	35	152	18.72	**1.89E-02**
BRCA1	5	630	0.79	6	181	3.21	**2.12E-02**
KCNT2	5	630	0.79	6	181	3.21	**2.12E-02**
PEG3	5	630	0.79	6	181	3.21	**2.12E-02**
FAT3	15	620	2.36	11	176	5.88	**2.24E-02**
FLG	20	615	3.15	13	174	6.95	**2.34E-02**
ASXL3	6	629	0.94	6	181	3.21	**3.44E-02**
GOLGB1	6	629	0.94	6	181	3.21	**3.44E-02**
PKHD1L1	11	624	1.73	8	179	4.28	**3.78E-02**
MAP1A	9	626	1.42	7	180	3.74	**4.99E-02**

Table 3: Genes with obvious deleterious mutations that are enriched based on ER status of tumor. Only obvious deleterious mutations (frame-shift, splice site, nonsense, nonstop) that occurred at least 5 times in ER+ patients and significantly enriched based on Fisher's exact test ($p<0.05$) are shown.

	ER+			ER-			
Hugo_symbol	Mutated	Not-mutated	%	Mutated	Not-mutated	%	**Fisher**
TP53	39	596	6.14	56	131	29.95	**3.05E-16**
GATA3	71	564	11.18	0	187	0	**4.24E-09**
MAP3K1	46	589	7.24	1	186	0.53	**5.72E-05**
CDH1	43	592	6.77	1	186	0.53	**1.22E-04**
MLL3	30	605	4.72	2	185	1.07	**1.25E-02**
MAP2K4	19	616	2.99	1	186	0.53	**3.78E-02**
NCOR1	18	617	2.83	1	186	0.53	**4.71E-02**

Task 3: Identification of genomic aberrations during breast cancer metastasis (months 6-36)

3a: Isolate DNA from tissues and perform library preparation (3 tissues from 10 individuals) (months 6-9)

HMW isolation, 5kb library prep, 40kb optimization, 3-12kb kit

Due to the nature of mate-pair sequencing, only frozen tissue can be utilized as FFPE severely degrade genetic material. Further, non-column based DNA extraction must be used to obtain high-molecular weight genomic DNA (gDNA). Pulse-field gel electrophoresis revealed high quality gDNA from frozen tumor tissue with a smear from ranging from ~30-200kb. A pilot library was run using Illumina Mate-Pair v2 library prep kit. Recently, Illumina updated this library prep kit to take advantage of the Nextera engineered retrotransposon technology. Several advantages were realized by this upgrade: (1) input DNA was reduced from 10ug to 4ug, (2) 'out-of-the-box' multiplexing compatible, and (3) multiple sized insert selections. The third point offers the ability to multiplex different sized mate pair library preps from the same input DNA from the same sample. Although larger insert sizes are more powerful at detecting structural rearrangements, library preparation is less efficient at higher insert sizes. By doing 3 independent library preps (3-5, 5-8, 8-12kb) for each sample allows these tradeoffs to be balanced. Each sample in the 'mate-pair' column of Table 1 was prepared in this way. The three libraries for each sample were multiplexed and run in a single lane on a HiSeq2000 (1 sample=3 libraries=1 lane).

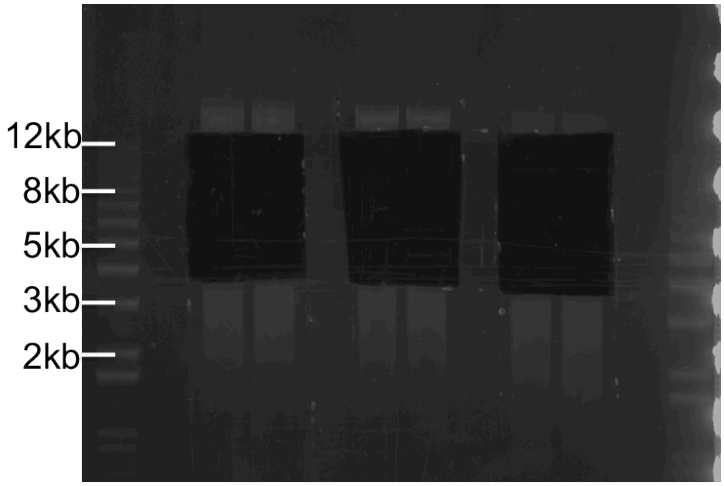

Figure 4: Size selection for Nextera Mate-Pair Library prep for RJH-MET-2. Each sample was sheared via random retrotransposon insertion and run in two lanes in a 0.6% megabase agarose gel for 2h at 100 V. The gel was stained via SYBR safe and visualized via a blue light transilluminator (to protect the DNA from UV damage). Three bands for each sample were extracted via clean razor blades for the remainder of the protocol. Post cutting, the gel was imaged with a traditional UV imager.

3b: Conduct sequencing (months 10-18)

As shown in Table 1, I have prepared mate-pair libraries and sequenced 7 samples from 2 individuals (RJH-MET-2, RJH-MET-3). In addition, a number of other analyses were performed on RJH-MET-3 allowing for deep integration of orthogonal dataset. Of particular note is the targeted amplification, bisulfite-seq analysis performed in collaboration with RainDance Technologies. This represents 1000 custom picked regions of interest (ROIs) of ~250bp with ~300x coverage. Many oncogene and tumor suppressor CpG islands are represented in these ROIs. Although analysis is still in the early stages, I have begun to integrate the findings from this dataset with the mate-pair sequencing data.

Each sample (3 library multiplex) produced ~160-180M 100bp pair-end reads, representing >225B base pairs of raw sequence. fastQC was run on all raw reads and revealed that mean read qualities were >30 for the entirety of both the forward and reverse reads. Extensive effort was placed in the development of an analysis pipeline, which is briefly described in Figure 5. All metrics were nominal throughout the library prep, sequencing, and analysis. Since I take the samples from tissue, to library prep, to sequencer, to analysis, and we have direct access to a HiSeq2000, I do not expect any changes in the quality, throughput, or timing for the remainder of the study. As seen in Figure 6, the insert sizes of the different libraries after alignment and removing duplicates have the expected distributions.

3c: Analysis of sequencing data and basic quality control (months 19-24)

The main motivation for sequencing the sample via mate-pair was to be able to detect structural variations. To visualize this data, circos software was used. Although I am still evaluating different breakpoint calling algorithms, the results from breakdancer integrated with the RDT methyl-seq and copynumber (readdepth) are show in Figure 7. The data shown (from outside in) is as follows: (1) chromosome ideogram with centromeres marked in red, (2) RDT methyl-seq (log2 ratio sample to normal; red>2, green<-2), (3) readdepth (log2 normalized to normal; red>2,

green<-2), (4) intra-chromosomal structural variants, and (5) inter-chromosomal structural variants (translocations) or large (>40Mb) intra-chromosomal structural variants.

Detailed analysis is ongoing but there are a few immediate observations that can be immediately observed from the circos plots.[4] First, almost all of the somatic methylation changes are *hyper*methylation. Next, there are a few, classic, arm level amplifications in the primary tumor (17q, 19q). These regions seem to also be hotspots for structural variants. Also, there seems to have been a major genetic event early during tumorigenesis involving a translocation bringing fragments of chromosome 1, 6, 7, 10, and 19 together in what looks like a 'daisy-chain' arrangement. Finally, it is also obvious that there are many structural changes that occur between the primary and metastatic disease. The 'quieter' nature of the local recurrence is potentially due to lower cellularity, but could also represent a less aggressive sub-clone of the original tumor that was not able to metastasize. Single nucleotide variant calling via GATK pipeline is underway and should allow more detailed analysis of clonal populations and tumor evolution.

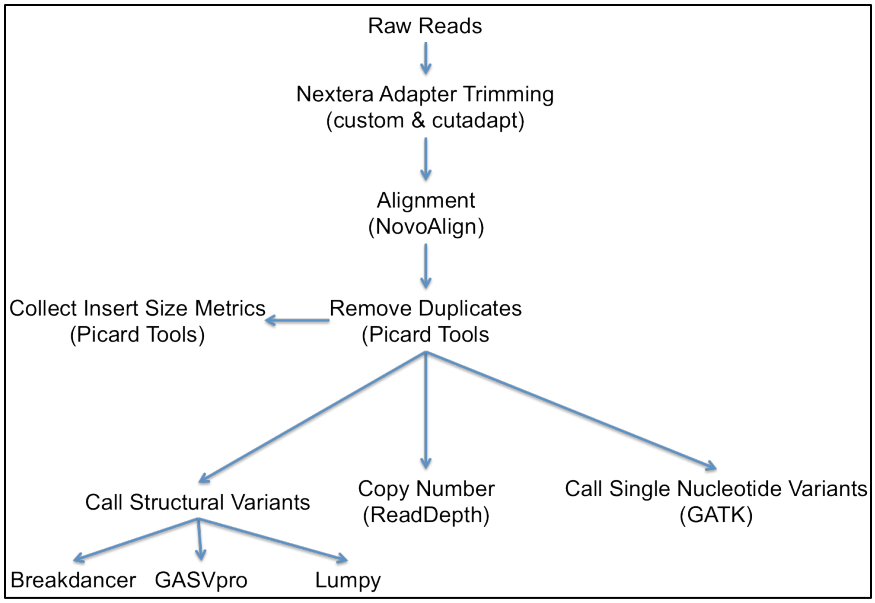

Figure 5: Mate-Pair Analysis Pipeline.

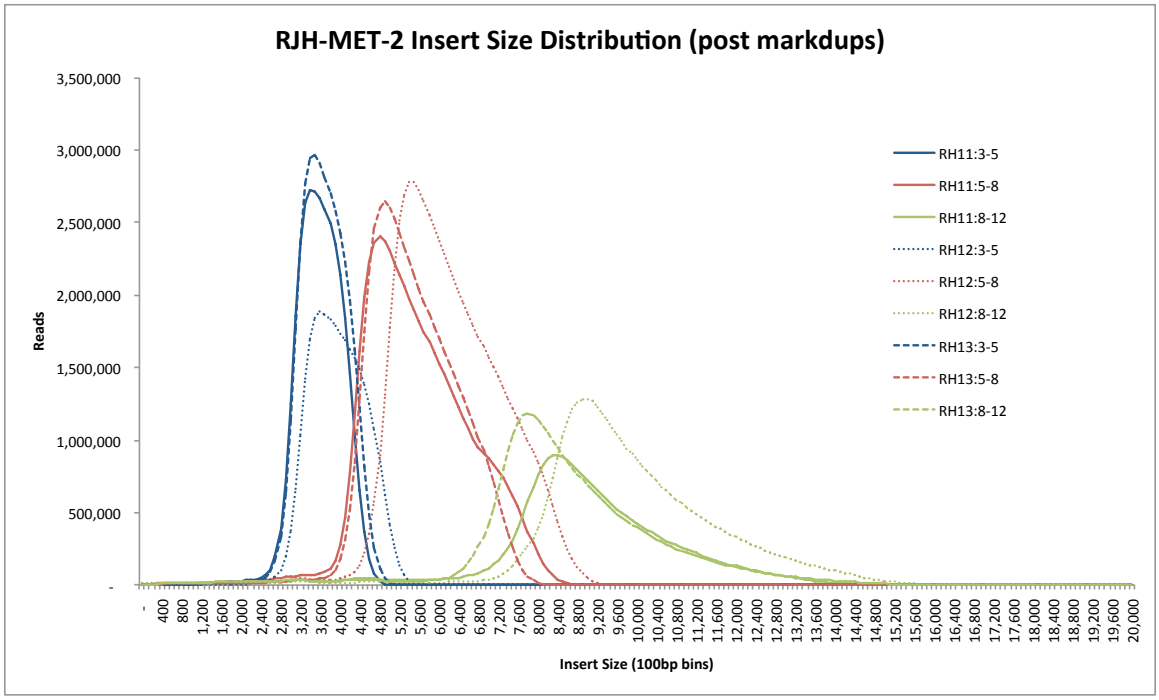

Figure 6: Post-alignment, post-duplicate removal insert size distribution for RJH-MET-2. Number of reads were counted in bins of 100bp.

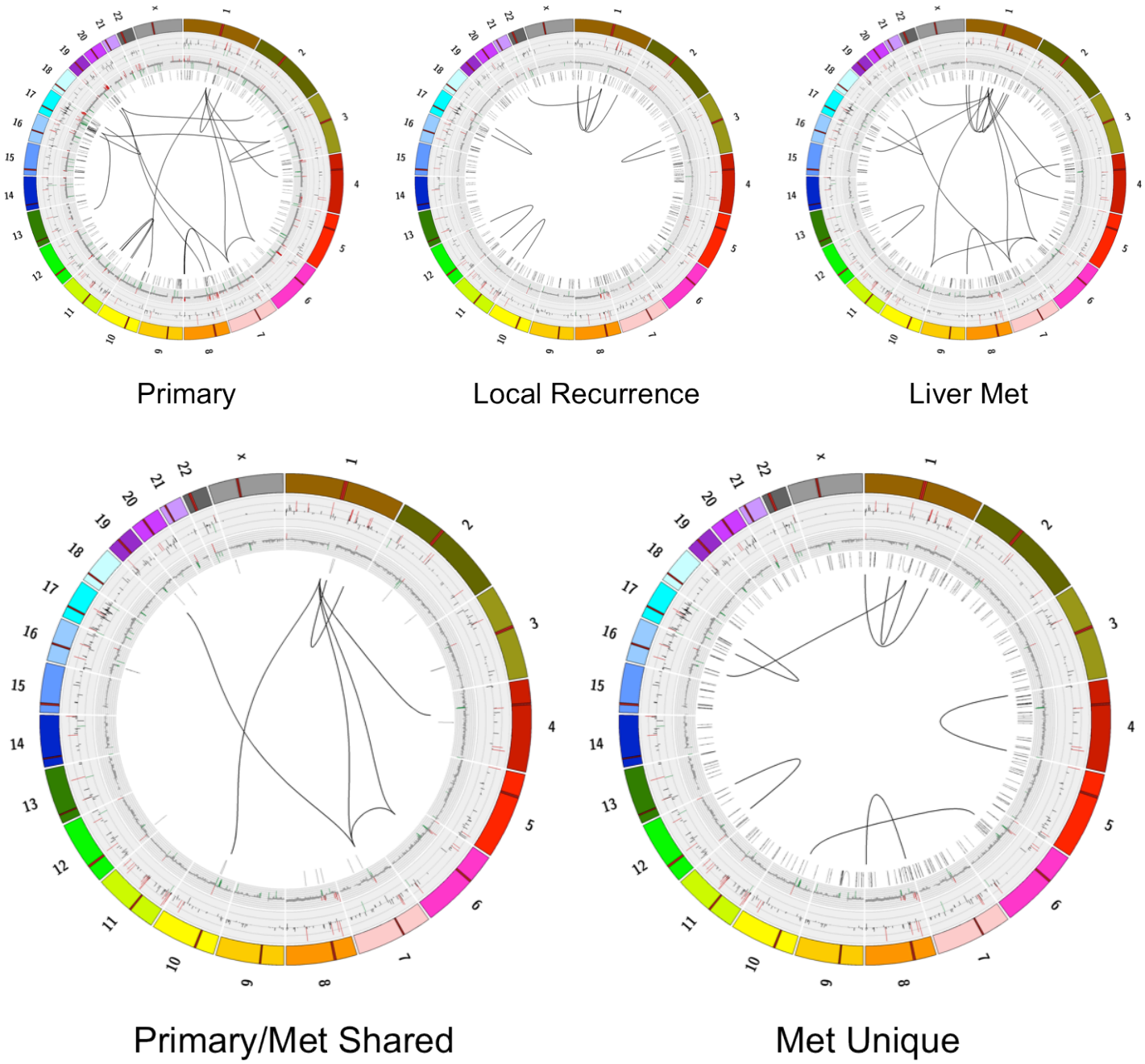

Figure 7: Circos plots of RJH-MET-3. Structural variations called by Breakdancer are visualized via links in the circos plots (inner 2 tracks). Other data track are as follows (outside to inside): (1) chromosome ideogram with centromeres marked in red, (2) RainDance Technologies target methyl-seq methylation levels (log2 ratio tumor:normal; red>2, green<-2); (3) copy number via readdepth (log2 ratio tumor:normal; red>2, green<-2), (4) intra-chromosomal structural variations, and (5) inter-chromosomal or large (>40Mb) intra-chromosome structural variations. Shared or unique structural variations were required to have both breakpoints within a 10kb window of the matching breakpoint.

3d: Systematic validation of identified rearrangements (months 25-30)

A major factor in being able to validate candidate rearrangements is the quality of the breakpoint calling. Thus, I am currently running two other callers (GASVpro, and Lumpy) on my samples. Both of these breakpoint callers utilize copy number, remapping, and split-read analysis to improve the power, specificity, and resolution of the breakpoint. All of this will aid in the efficient validation of structural variations.

3c: Functional studies of selected rearrangements (months 31-36)

Not started

Key Research Accomplishments

- Obtained matched normal-primary-metastatic tissue pairs
- Unbiased identification of candidate driver genes in endocrine resistance, including NCOR1, through mining publically available datasets
- Integration of multiple orthogonal datasets
- Multiple insert size mate pair library prep and sequencing of 7 samples from 2 patients
- Analysis pipeline developed and customs scripts automated running
- Identification of primary/met shared and metastasis specific structural rearrangements

Reportable Outcomes

- Abstracts/Poster presentations
    - 2012.06: University of Pittsburgh Cancer Institute retreat
    - 2012.09: University of Pittsburgh Women's Cancer Research Center retreat
    - 2012.10: AACR Advances in Breast Cancer Research (upcoming)
- 2012.10.12 – 2012.10.30: Cold Spring Harbor Laboratory Programming for Biologists course
- Mate-pair sequencing and other data is being used as preliminary data for a DoD BCRP Breakthrough Award by my mentor
- Publications
    - **Hartmaier RJ**, Priedigkeit N, Lee AV. Who's driving anyway? Herculean efforts to identify the drivers of breast cancer. Breast Cancer Res. 2012 Oct 31;14(5):323
    - Smith CL, Migliaccio I, Chaubal V, Wu MF, Pace MC, **Hartmaier R**, Jiang S, Edwards DP, Gutiérrez MC, Hilsenbeck SG, Oesterreich S. Elevated nuclear expression of the SMRT corepressor in breast cancer is associated with earlier tumor recurrence. Breast Cancer Res Treat. 2012 Nov;136(1):253-65
    - Coronnello C, **Hartmaier R**, Arora A, Huleihel L, Pandit KV, Bais AS, Butterworth M, Kaminski N, Stormo GD, Oesterreich S, Benos PV. Novel modeling of combinatorial miRNA targeting identifies SNP with potential role in bone density. PLoS Comput Biol. 2012;8(12)
    - Osmanbeyoglu HU, **Hartmaier RJ**, Oesterreich S, Lu X. Improving ChIP-seq peak-calling for functional co-regulator binding by integrating multiple sources of biological information. BMC Genomics. 2012;13 Suppl 1:S1.

Conclusion

Although there are many lines of evidence pointing to a copy number variation in NCOR2, I have been unable to reliably validate it. Recent access to new equipment may help alleviate this problem. Regardless, based on recently published data, it is safe to conclude that NCOR2 does not *acquire* somatic CNVs during breast carcinogenesis at an appreciable frequency. NCOR1, on the other hand, has been shown to be mutated at ~2-3% at appreciable frequencies in all breast tumors. I have shown that NCOR1 as well as a number of other genes is significantly enriched in ER+ disease. Of note, GATA3 is exclusively mutated in ER+ disease making it likely that GAT3 plays a critical role in the underlying biology of ER+ disease. High quality genomic DNA has been isolated from matched normal, primary, and metastatic breast cancers and has subsequently been used to generate mate-pair libraries and sequenced. Additional technologies are being applied and the integration of orthogonal datasets is producing novel insights into the mechanisms underlying breast cancer metastasis.

References

1. Kidd, J. *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* **453,** 56–64 (2008).
2. Zhang, L. *et al.* SpliceArray Profiling of Breast Cancer Reveals a Novel Variant of NCOR2/SMRT That Is Associated with Tamoxifen Resistance and Control of ER Transcriptional Activity. *Cancer Research* **73,** 246–255 (2013).
3. Stephens, P. *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462,** 1005–1010 (2009).
4. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Research* **19,** 1639–1645 (2009).

Appendices

None

Supporting Data

None