

# Microsatellite Polymorphisms in Human Disease: Exploring ERR- $\gamma$ as a Potential Breast Cancer Biomarker

Sara M. Linehan, C. L. Galindo, J. McCormick, L. McIver, L. Shan, A. Rommel, D. Boothman, and H. R. Garner  
UT Southwestern Medical Center, Dallas, TX 75390

## Introduction

Microsatellites are typically defined as tandemly repeated sequences (motifs) of one to six nucleotides that are very widely distributed throughout the genome (~500,000 microsatellite-containing loci) and are frequently variable in the number of times the motif is repeated (vary 30-50% among individuals). Microsatellite expansions are causative for over 20 neurological diseases, and microsatellite alterations occur in most tumors, with certain types of tumors (e.g., hereditary non-polyposis colorectal cancers) harboring significantly elevated rates of microsatellite mutations. Alterations in repeat unit number in and around coding sequences can have important quantitative and qualitative effects on gene expression, with a growing body of evidence to support microsatellite length correlating with various phenotypes (e.g., circadian rhythm in fruit flies, snout length in dogs, social behavior in voles and possibly humans). Nonetheless, microsatellites are underappreciated and understudied, mainly due to the difficulty in accurately sequencing them and lack of method to measure/examine them *en masse*. Our laboratory developed both a custom array and computational methods to circumvent these limitations, and we are finding that microsatellite variability is more pervasive and linked with many more diseases than previously thought. This study highlights the methodology used to search for polymorphic microsatellites of consequence and subsequent identification and characterization of new biomarkers.

## Experimental Approach

**Custom Global Microsatellite Microarray genomic summated microsatellite content**

	All Genes	Cancer Genes	Neurological Genes
Upstream	131,180	864	305
Downstream	3,629	710	270
5' UTR	5,411	717	260
Intron	168,319	38,464	20,944
Exon	1,124	296	161
Intergenic	289,705		
<b>Total</b>	<b>507,590</b>	<b>47,134</b>	<b>24,334</b>

**Neurological Diseases**

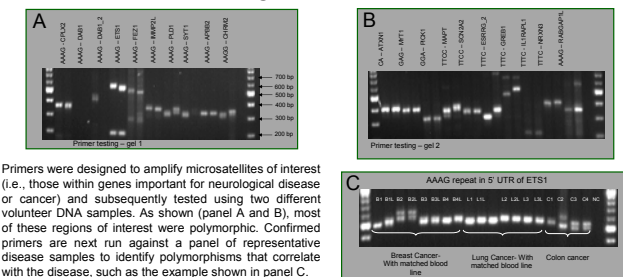
	Total in genome	In or near genes	Differ in length (humans vs chimps)	Promoter regions
500,000	207,885	98,601	14,476	

There are ~500,000 microsatellites (at least 18-20 bp long) in the human genome. In order to study them systematically, we prioritized them based on location (e.g., promoter regions), likelihood of being polymorphic (e.g. shorter motifs, longer sequence), and their likelihood of contributing to/correlating with human disease (e.g., located in genes known or suspected to be involved in cancer). Shown in the tables above are generalized schemes for narrowing down the search for cancer and neurological disease biomarkers.

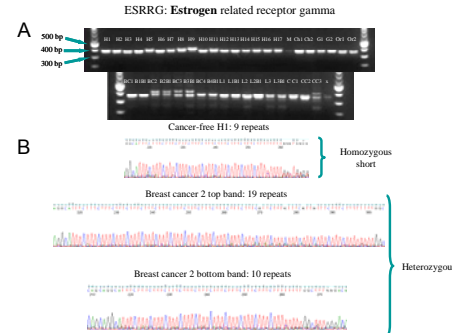
**Length-based PCR design primers, test on target disease panels, sequence confirmation**

**Functional Analyses**  
gene expression, knockdown experiments, Western blots, luciferase reporter assays

## Microsatellite Polymorphisms identified by Length-based PCR



## Polymorphic (AAAG)<sub>n</sub> in 5'UTR of ERR- $\gamma$



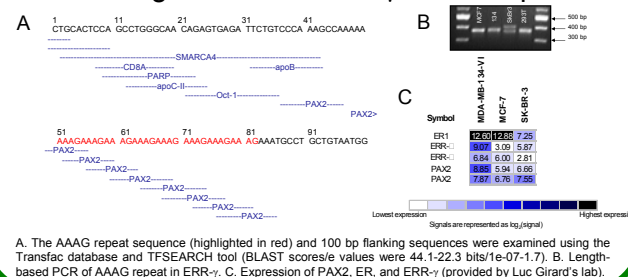
A quick gel survey of the ERR- $\gamma$  locus (A) followed by sequencing (B) are shown. All cancer free humans samples (H1-17) possess 7-10 tandem copies of AAAG within the 5' UTR of the ERR- $\gamma$  gene (18q21.2), while breast cancer 2 and 3 (BC2 and BC3, HCC2157 and HCC1187 cell lines, respectively) with their matched blood lines (B2BI, B3BI), as well as colorectal cancer 3 (CC3, RKO cell line) are heterozygous at the loci, with upper bands ranging from 19-21 repeats. To validate polymorphism specificity in human disease, a series of animal controls were also used: M = mouse, Ch = chimpanzee, G = gorilla and O = orangutan.

## The AAAG repeat in ERR- $\gamma$ is a potential marker for breast cancer predisposition

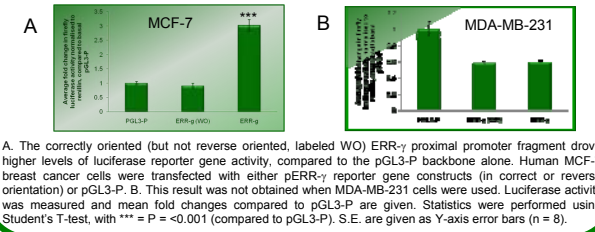
	Non-carriers		Carriers		Totals		Incidence		Statistics (p value)	
	Healthy: no BC family hx	Healthy: all	Healthy: no BC family hx	Healthy: all	Healthy: no BC family hx	Healthy: all	Healthy: no BC family hx	Healthy: all	Healthy: no BC family hx	Healthy: all
Healthy volunteers:										
No BC family hx	100	5	105	4.8%	n = 105	n = 174				
BC family hx	64	69	133	8.2%	0.2992	0.5143				
Cancer patients:										
Breast cancer	126	21	147	14.35%	0.0195	0.0130				
Colorectal cancer	45	6	51	11.8%	0.1785	0.2100				
Other sample types:										
Colorectal polyps	48	5	53	9.4%	0.3049	0.3504				
Lung cancer cell lines	21	1	22	4.5%	1.0000	1.0000				
Totals	404	43	447	9.6%	0.1262	0.1498				

Based on genotyping of 447 cancer-free volunteers and cancer patients (germlines), the size of the AAAG motif ranged between 5 and 21 copies. Carriers (13+ copies) and non-carriers (less than 13 copies) of the longer allele for each category of patient are presented. As shown, a statistically significant higher incidence of long allele carriers (p value = 0.0195, two tailed Fisher's exact test) was observed for breast cancer patients (14.3%), compared to healthy volunteers (4.8%), which translates to a relative risk ratio of 2.98 (14.3/4.8).

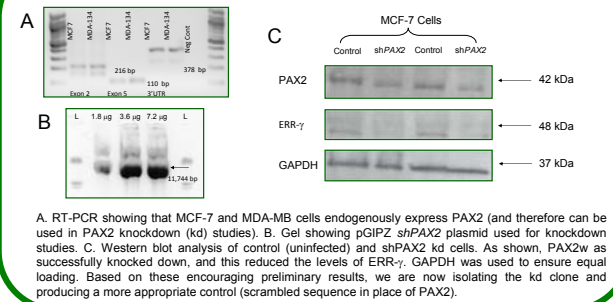
## PAX2 might bind the ERR- $\gamma$ AAAG repeat



## ERR- $\gamma$ promoter drives Luciferase activity



## PAX2 is required for ERR- $\gamma$ expression



## Conclusions and Future Directions

- Based on initial length-based PCR results, there are many more polymorphic microsatellites within genes important for neurological development than we initially anticipated – identification of new biomarkers for neurological diseases is thus promising.
- Our high-throughput methodology for studying microsatellites en masse has led to the discovery of at least one potential biomarker – a repeat in the 5'UTR of ERR- $\gamma$  that appears to correlate with breast cancer predisposition.
- Initial experiments indicates that PAX2 can bind to the AAAG repeat in the 5'UTR of ERR- $\gamma$  and drive its transcription. Indeed, knockdown of PAX2 reduces levels of ERR- $\gamma$  in MCF-7 breast cancer cells.
- Future work will include investigation of the PAX2-ERR- $\gamma$  interaction in the context of other hormone markers (ER, PR, HER2) and tamoxifen resistance. The Garner lab will continue to search for new microsatellite biomarkers.

## Acknowledgements

I would like to thank Eastfield College, the National Science Foundation Grant # DUE-0525536, UT Southwestern, and the STARS Program for giving me this great research experience at one of the top medical centers in the country. Thank you also to Laura Thomason, Norma Mendoza, Dr. Carl Knight and the Garner lab for their guidance, helpful comments, and discussions throughout the summer.