The role of repetitive elements in ovarian cancer initiation and progression

Uzma Rentia, Tomas Kanholm, Katherine Chiappinelli

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Abstract

Background: Ovarian cancer is currently the most lethal gynecologic malignancy. Despite the poor prognosis, most ovarian cancers are detected at Stages III and IV due to the lack of early diagnostic methods and ambiguous disease presentation. In recent years, repetitive elements (REs) have emerged as a potential early diagnostic marker for ovarian cancers. While expression of REs is largely silenced in normal cells via methylation, they are often transcribed in cancer, likely due to broad hypomethylation of cancer genomes. Although broad overexpression of REs has been observed in ovarian cancer, specific REs that are consistently upregulated in cancer progression have not been identified, nor have their impacts on disease progression been evaluated. In addition, there has not been sufficient research analyzing differential RE expression and methylation in tandem to discern if hypomethylation truly causes RE expression during cancer initiation. The aim of this study was to identify significantly upregulated or downregulated REs and their associated methylation profiles in both immortalized fallopian tube secretory epithelial cells (FTSEC) and a cancer progression model.

Results and Conclusions: The cancer progression model showed differential expression compared to controls. Most elements were overexpressed as predicted, though there were a significant number of repetitive elements demonstrating reduced expression. The methylation analysis showed evidence of broad hypomethylation. Unlike hypothesized however, the immortalized FTSECs had near ubiquitous downregulation of repetitive elements.

Methods

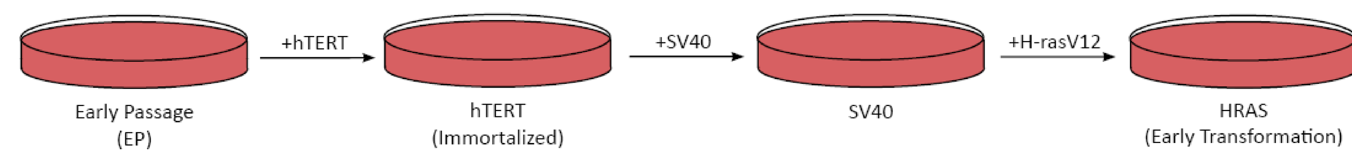

Figure 1. Stepwise transduction of BJ fibroblasts to generate cancer progression model done by Easwaran Lab

A. BJ fibroblast cells underwent serial transduction with driver mutations hTERT, SV40 large T antigen, and HRAS to create a cancer progression model
B. RNA-Seq data and whole genome bisulfite sequencing data was collected from the cells at each stage of progression.
C. RNA-Seq data underwent underwent quality controls using the softwares FastQC and Cutadapt. The cleaned data was then aligned to the reference genome hg38 using the software STAR.
D. Read calling of repetitive elements (REs) was done concurrently on two platforms, TEtranscripts and Telescope, after which differential expression analysis was done on resultant count tables using DESeq2.
E. Methylation analysis was conducted on methylKit using whole genome bisulfite sequencing data.

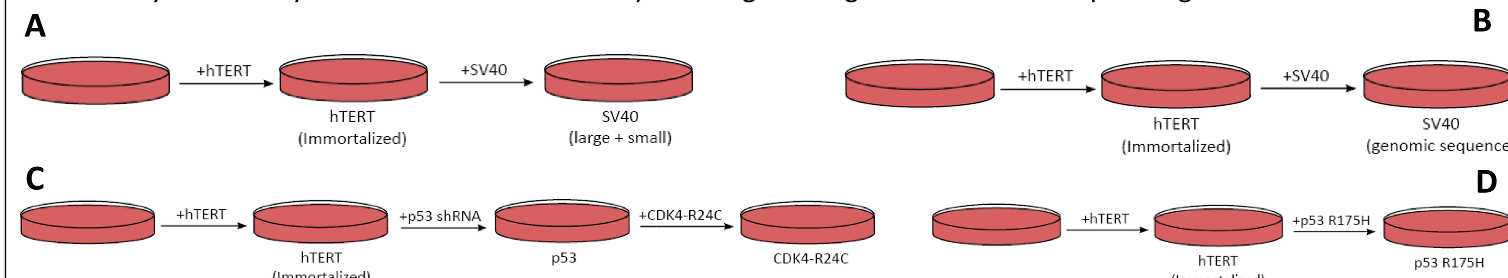

Figure 2. Stepwise alterations to create immortalized FTSEC cells done by Drapkin Lab. Figure 2A depicts the steps to create cell line FT33-TAG; Figure 2B depicts the steps to create cell lines FT189, FT190, and FT194; Figure 2C depicts the steps to create cell lines FT237, FT240 and FT246; and Figure 2D depicts the steps to create cell line FT282. The multiple cell lines can better model ovarian cancers with different mutational profiles.

F. After preliminary results were found using the cancer progression model, RNA-Seq data was collected from immortalized fallopian tube secretory epithelial cells (FTSEC). RNA-Seq data was also collected from the fallopian tissue of control patients.
G. RNA-Seq data underwent the same quality control as the immortalized BJ fibroblast cell line, followed by removal of potential rRNA contamination with SortMeRNA.

RNA-Seq Analysis of BJ Fibroblast Data

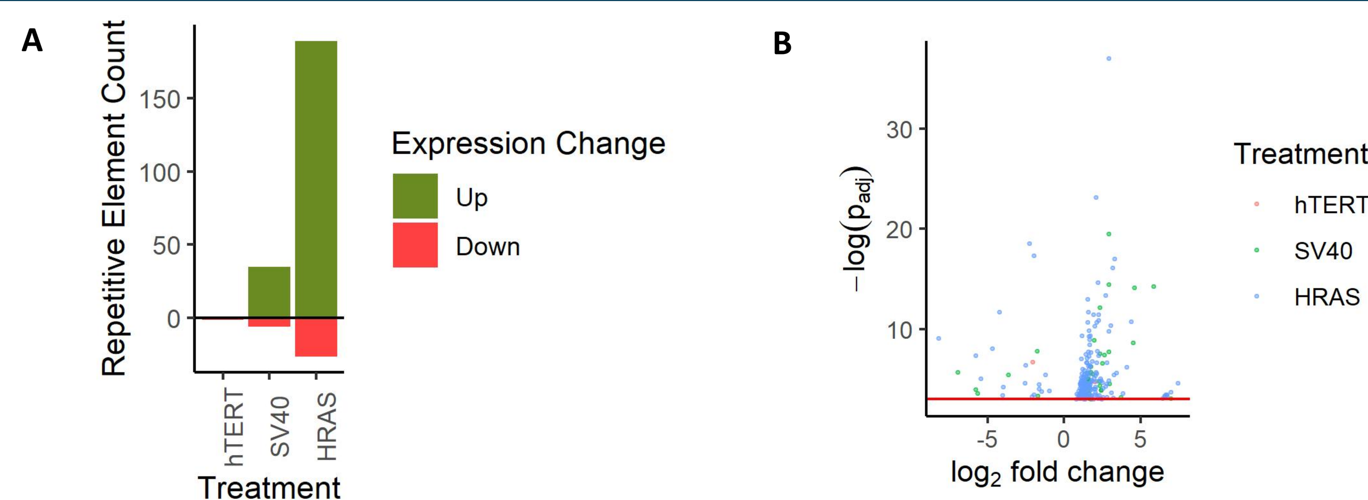

Figure 3. Summary results of TEtranscripts analysis. The software TEtranscripts is notable for its ability map short sequencing reads to repetitive regions of the genome. It can also handle multi-reads, or short reads that may map to multiple sites of the reference genome. These elements are often attributed to transposable elements. While many analysis tools ignore them or have poor accuracy given the high-level ambiguity, TEtranscripts can handle repetitive elements well in addition to conventional genes, making it a good candidate to assist with differential expression analysis. As hypothesized, the overexpression of REs increased at each step of the transformation process (Figure 3A). The volcano plot (Figure 3B) indicates that most of the differentially expressed elements had a P-value of < 0.05. Figure by Tomas Kanolm.

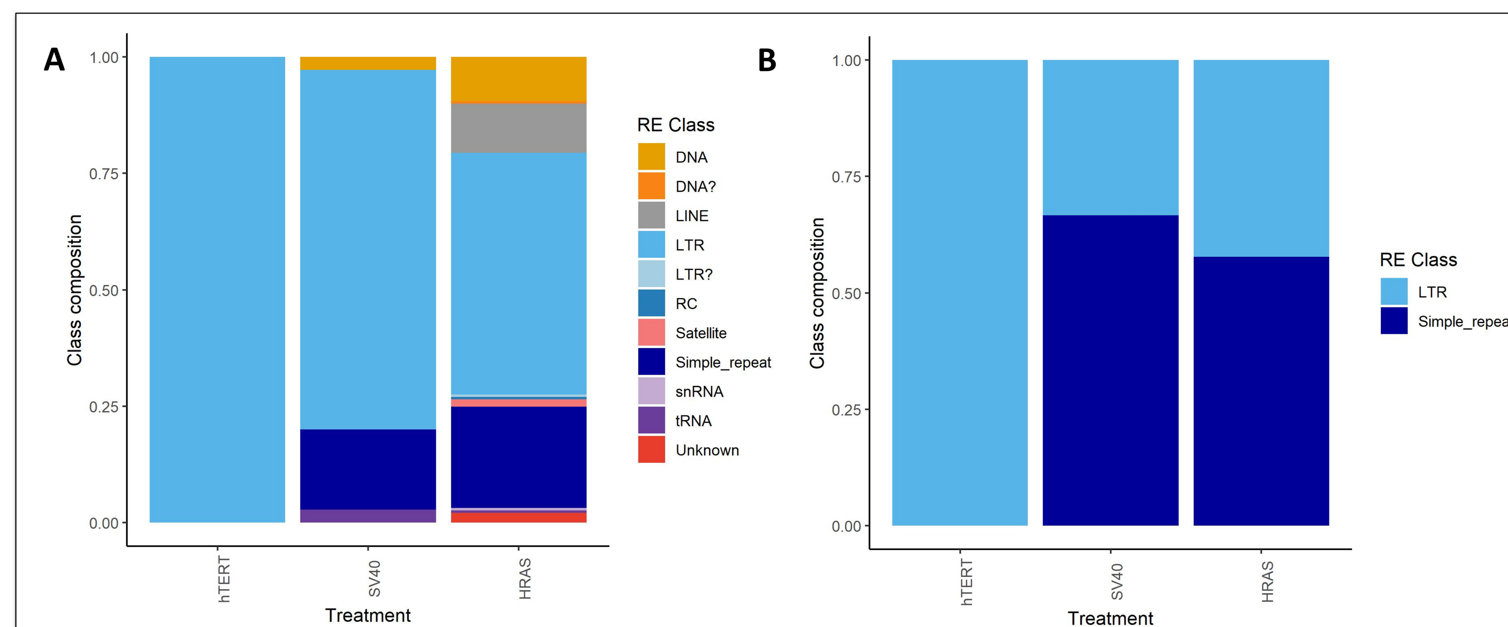

Figure 4. Class composition of differentially expressed elements. Differential expression was determined by comparing each step of progression to the to early passage (EP) stage. Long Terminal Repeats (LTR) are the most commonly over- (Figure 4A) and underexpressed (Figure 4B) RE. This is significant because LTR elements can function as both enhancers and promoters, so their errant expression could impact the transcription of protein-coding genes or other REs. The introduction of the HRAS mutation also saw an increase in expression of long interspersed elements (LINEs). Long interspersed element-1 comprises 17% of the human genome and some have retained autonomous retrotransposition capabilities. These elements are potentially responsible for large deletion events that remove tumor suppressor genes. Figure by Tomas Kanholm.

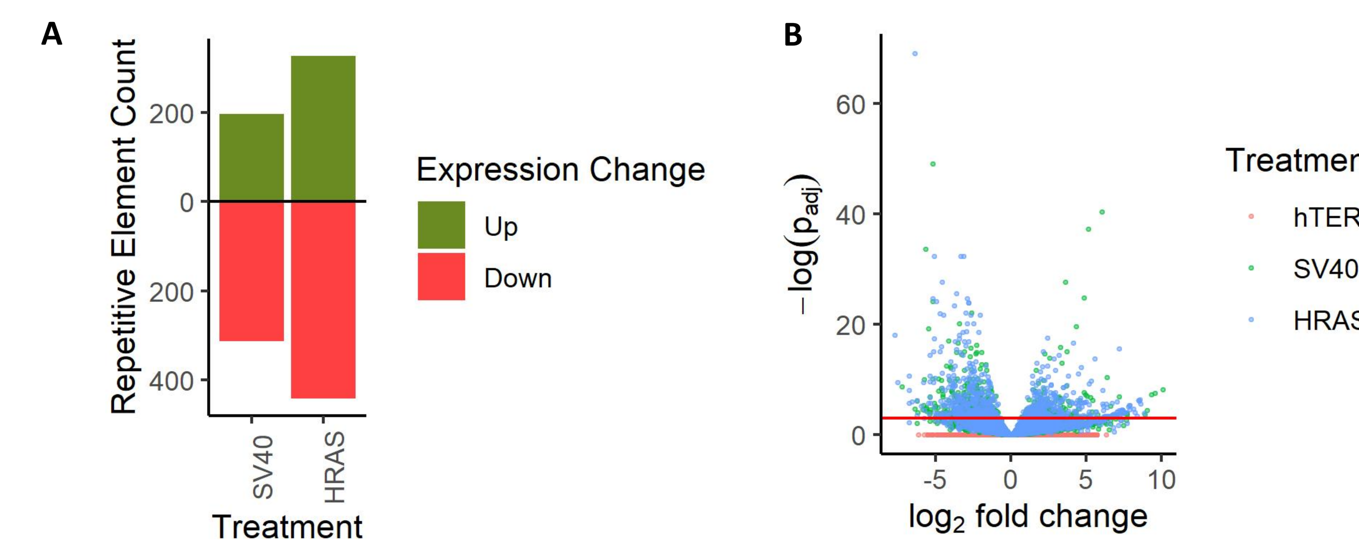

Figure 5. Summary results of Telescope analysis. Telescope, created by the Crandall Lab at the GW Computational Biology Institute, is distinct from TEtranscripts for its ability to provides locus-specific quantification of select REs. While the plots indicate increased dysregulation of expression as mutations are introduced as expected, far more elements are downregulated compared to the TEtranscripts results. This could be due to the fact that Telescope examines a more limited range of REs than TEtranscripts. Figure by Tomas Kanholm.

Methylation Analysis of BJ Fibroblast Data

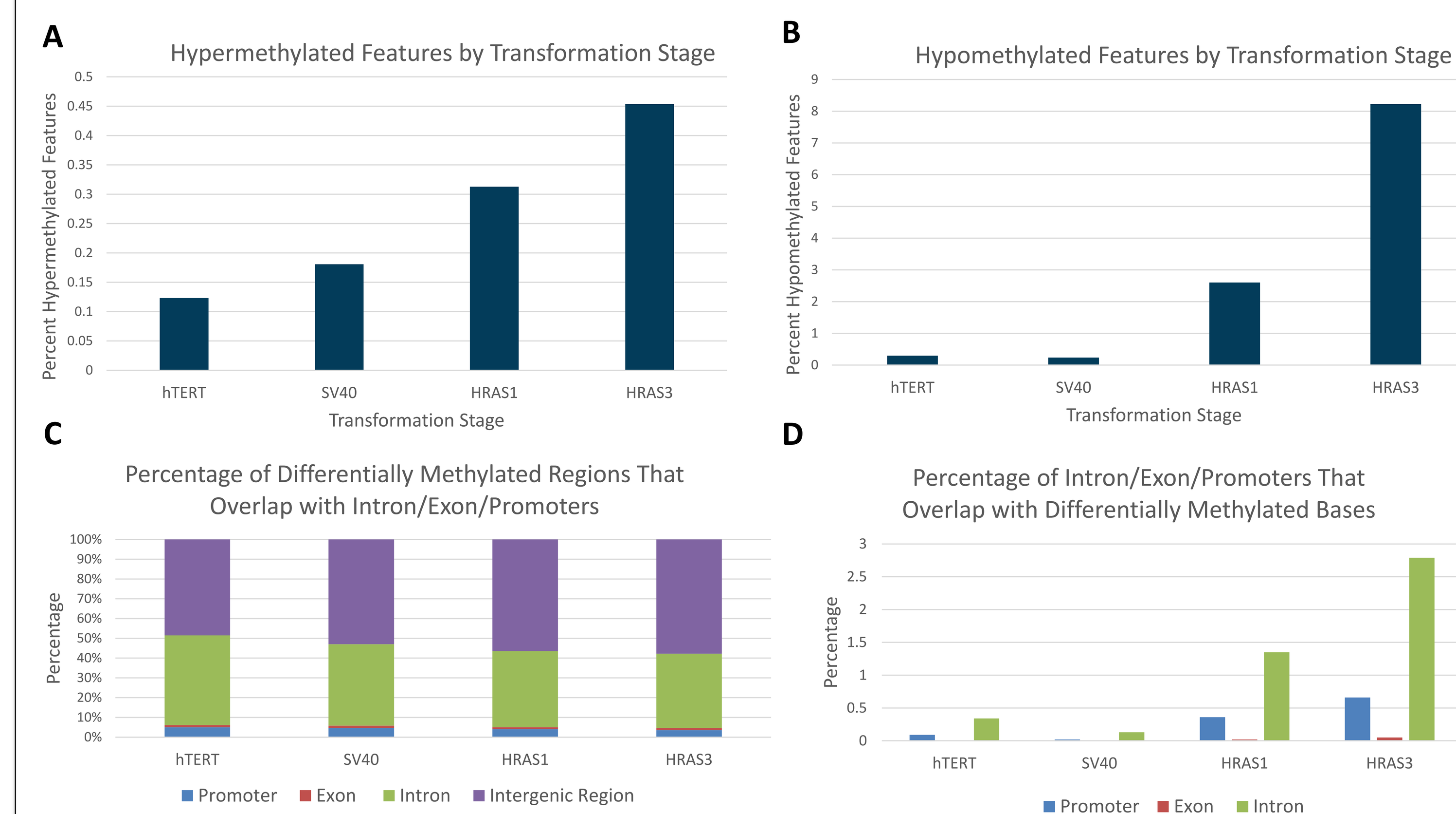

Figure 6. Summary results of methylKit methylation analysis. These results indicate that there was some level of hypermethylation (Figure 6A), but the overall difference is much smaller than hypomethylation. As predicted, the percentage of hypomethylated features increased as the BJ fibroblast cells were immortalized (Figure 6B). Despite HRAS1 and HRAS3 being replicates, the two tests showed large differences in results. For example, 8.22% of HRAS3 features were hypomethylated while only 2.6% of HRAS1 features were. At the final stage of the immortalization process—the introduction of HRAS—0.66% of promoters and 0.05% of exons overlapped with differentially methylated regions (Figure 6D). Most DMRs overlapped with introns and intergenic regions, which is to be expected given that they occupy a large portion of the entire genome.

RNA-seq Analysis of Immortalized FTSEC Data

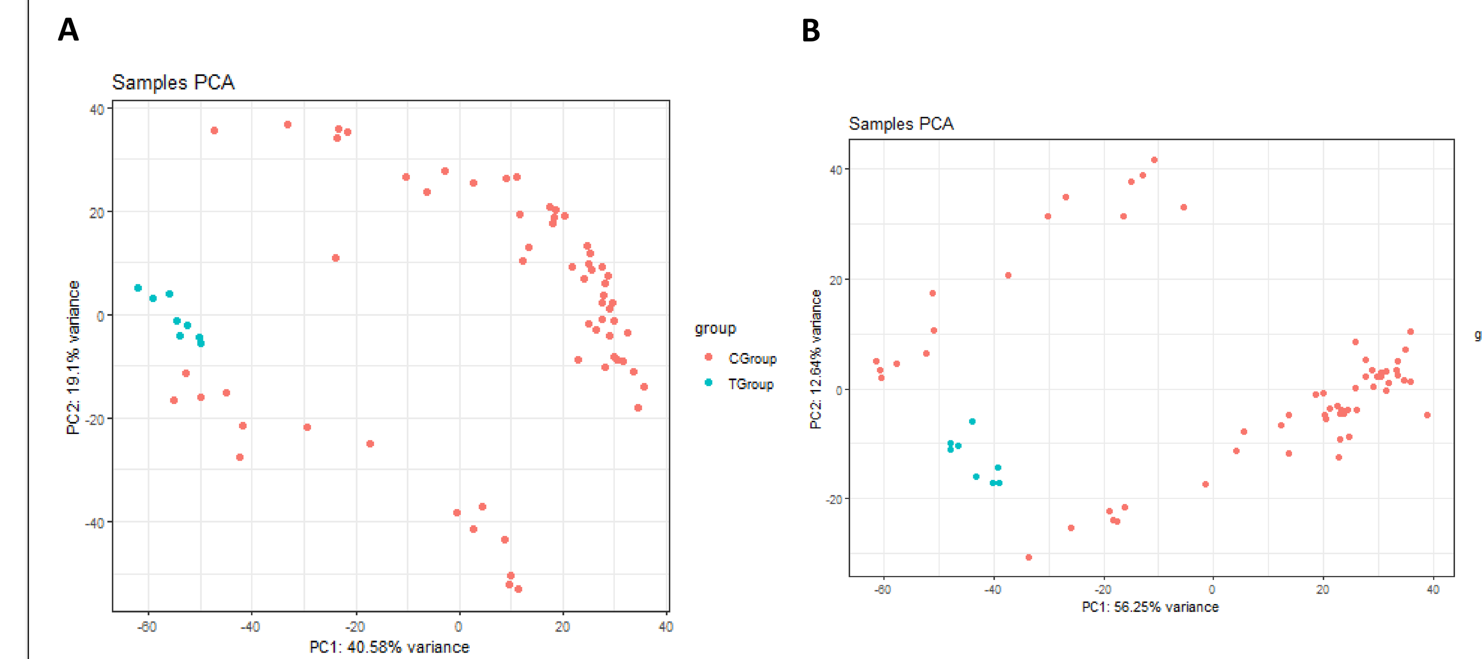

Figure 7. Prinical Component Analysis (PCA) of immortalized FTSECs and normal fallopian tube epithelium (NFTE). Principal component analysis of the data reveals the formation of two independent clusters within the normal fallopian tube epithelium (NFTE) cells in both the TEtranscripts (Figure 7A) and Telescope (Figure 7B) results. However, the immortalized cells aggregate close together in both plots. Analysis was done following the removal of potential in silico rRNA contamination.

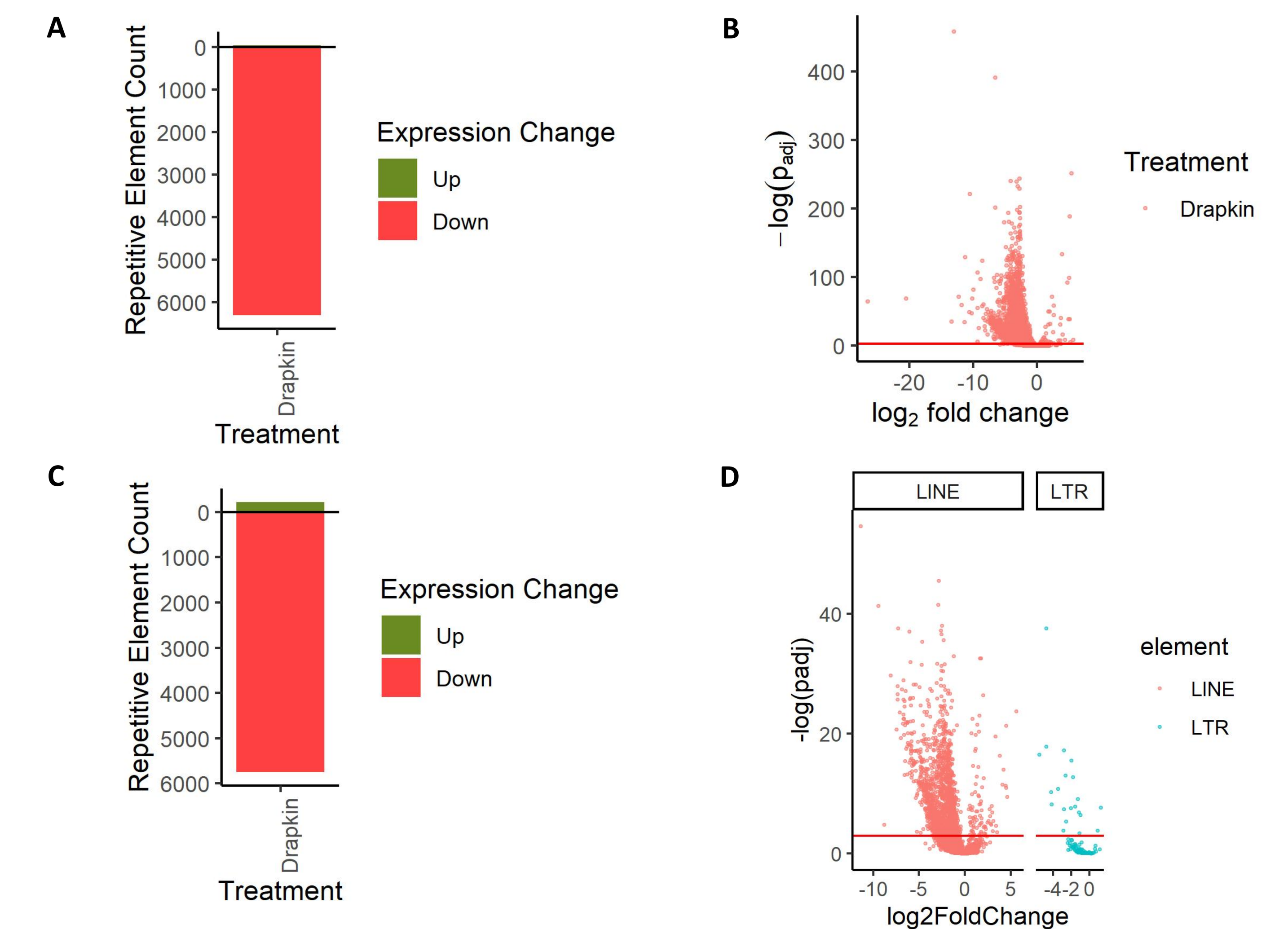

Figure 8. Summary results of TEtranscripts and Telescope analysis. Unlike the transformed BJ Fibroblast lines, the immortalized FTSECs show widespread downregulation of repetitive elements compared to NFTE in both TEtranscripts (Figure 8A) and Telescope (Figure 8C). The Telescope results show that while many LINE and LTR elements show statistically significant differential expression, they are largely underexpressed compared to NFTE cells (Figure 8D). Figure by Tomas Kanholm.

Conclusion and Acknowledgement

A. The cancer progression model showed both overexpression and hypomethylation, results supported by previous studies. Moving forward, repetitive elements that are substantially and significantly differentially expressed will be mutated in ovarian cancer models to determine phenotypic changes.
B. While methylKit provides good summary statistics, other softwares provide specific genomic coordinates to differentially methylated regions (DMRs). Future directions include using DSS and WIMSi to isolate DMRs for further study.
C. Immortalized fallopian tube secretory epithelial cells (FTSEC) were found to induce decreased expression of elements compared to controls. This result was unexpected given previous findings. Future work includes determining the reasons behind this change and the cause of the clustering seen on the PCA plots.

This work was supported by:
A. Katherine Chiappinelli, Tomas Kanholm, Stephanie Gomez, James McDonald, Erin Grundy, and Melissa Beaty
B. Ronny Drapkin's lab at the Perelman School of Medicine for providing the immortalized FTSECs
C. Hari Easwaran's lab at John Hopkins for transforming and sequencing the BJ fibroblast cells
D. Kate Lawrenson's lab at Cedars-Sinai Medical Center for providing the NFTE data