Mutations in Noncoding Cis-Regulatory Elements Reveal Cancer Driver Cistromes in Luminal Breast Cancer

Samah El Ghamrasni1, Rene Quevedo1,2, James Hawley1,2, Parisa Mazrooei1,2,3, Youstina Hanna1, Iulia Cirlan1, Helen Zhu1,2,4,5, Jeff P. Bruce1, Leslie E. Oldfield1, S. Y. Cindy Yang1,2, Paul Guilhamon6,7, Jüri Reimand2,5,8, Dave W. Cescon1, Susan J. Done1,2,9, Mathieu Lupien1,2,5, and Trevor J. Pugh1,2,5

ABSTRACT

Whole-genome sequencing of primary breast tumors enabled the identification of cancer driver genes and noncoding cancer driver plexuses from somatic mutations. However, differentiating driver from passenger events among noncoding genetic variants remains a challenge. Herein, we reveal cancer-driver cis-regulatory elements linked to transcription factors previously shown to be involved in development of luminal breast cancers by defining a tumor-enriched catalogue of approximately 100,000 unique cis-regulatory elements from 26 primary luminal estrogen receptor (ER)+ progesterone receptor (PR), and HER2/NEU (HER2; ref. 1). Approximately 80% of all breast cancers are of the luminal (ER+) subtype with luminal B (ER+/PR−/HER2+) being the most proliferative and aggressive and luminal A (ER+/PR+/HER2−) being the most predominant. Sixty-five percent of luminal breast tumors are ER+ PR−; together, this luminal subtype makes up 52% of all breast cancers (2, 3). Large-scale analysis of whole-genome sequencing (WGS) in breast tumors has identified 99 driver genes with recurrent protein-coding alterations (4, 5) as well as a high number of mutations within the noncoding genome (4). Noncoding mutations can alter the transcription factor binding to the DNA and affect enhancer–promoter interactions to perturb gene expression (6–12). However, the inclusion of noncoding mutations to find cancer drivers remains a challenge in ER+ PR− luminal breast cancer that needs to be addressed to comprehensively resolve the role of genetic variants in oncogenesis.

Implications: Mapping the accessible chromatin of luminal breast cancer led to discovery of an accumulation of mutations within cistromes of transcription factors essential to luminal breast cancer. This demonstrates coopting of regulatory networks to drive cancer and provides a framework to derive insight into the noncoding space of cancer.

Introduction

Breast cancer is the second leading cause of death in women in North America (1). Currently, treatment decisions rely on the histology and the expression of three proteins: estrogen receptor (ER) α, progesterone receptor (PR), and HER2/NEU (HER2; ref. 1). Approximately 80% of all breast cancers are of the luminal (ER+) subtype with luminal B (ER+/PR−/HER2+) being the most proliferative and aggressive and luminal A (ER+/PR+/HER2−) being the most predominant. Sixty-five percent of luminal breast tumors are ER+ PR−; together, this luminal subtype makes up 52% of all breast cancers (2, 3). Large-scale analysis of whole-genome sequencing (WGS) in breast tumors has identified 99 driver genes with recurrent protein-coding alterations (4, 5) as well as a high number of mutations within the noncoding genome (4). Noncoding mutations can alter the transcription factor binding to the DNA and affect enhancer–promoter interactions to perturb gene expression (6–12). However, the inclusion of noncoding mutations to find cancer drivers remains a challenge in ER+ PR− luminal breast cancer that needs to be addressed to comprehensively resolve the role of genetic variants in oncogenesis.

The noncoding genome is known to harbor many of cis-regulatory elements, defined as binding sites for transcription factors involved in transcriptional regulation by serving as promoters, enhancers, or anchors of chromatin interactions (13). In luminal breast cancer, cis-regulatory elements are bound by key transcription factors, including ER, FOXA1, and GATA3 which have a role in maintaining the luminal phenotype as well as the growth and differentiation of breast epithelium (14). Disruption of either of these transcription factors or their binding sites can affect their binding to the chromatin (15), which can modulate downstream gene expression. A subset of transcription factors active in luminal breast cancer are classified as driver genes due to an enrichment of mutations within protein coding regions (16–18). Mutations within regulatory elements of enhancers and promoters can be responsible for the development of disorders with the same magnitude as mutations affecting protein-coding genes (6, 19–24). A classic example of this is the TERT promoter which is frequently mutated across several cancer types as a mechanism for telomerase reactivation (25); it has been observed in 71% of sporadic melanoma and 60% to 75% of glioblastomas (6, 19–24). Variants within the TERT promoter also lead to an increased risk of breast and ovarian cancer development (26). Analysis of the Pan-Cancer Analysis of Whole Genomes (PCAWG) project showed that the long tail of infrequent noncoding mutations in promoters and distal regulatory elements converged to pathways and molecular interaction networks of oncogenic processes (10). Zhule and colleagues found frequently mutated...
regulatory elements in cancer genomes that interact with target genes via long-range chromatin interactions (10).

The sum of all regulatory elements bound by a transcription factor in a given cell type has been referred to as a "cistrome" (27). Analysis of mutations across the cistromes of prostate cancer revealed a high frequency of mutations within the binding sites of key transcription factors including FOXA1, HOXB13, and androgen receptor (AR; ref. 7). In luminal breast cancer, Bailey and colleagues found 7 functionally validated mutations within the cis-regulatory elements of ESRI that altered gene expression (8), while Cowper-Sallari and colleagues found that risk-associated SNPs in the cistrome of FOXA1 modulated the expression of downstream target genes (15). These studies highlight the key, albeit underappreciated role that cis-regulatory elements and cistromes play in tumorigenesis.

Within this study, we drew parallels to the approaches to finding driver mutations between the coding and noncoding genome. In keeping with the terminology established by The Cancer Genome Atlas (TCGA) and the PCAWG, we define cancer drivers as units of the genome that are enriched in mutations more than expected by chance (5, 28). Similar to looking for hotspots of mutations within the genome that are enriched in mutations more than expected by chance, we first focused on individual cis-regulatory elements across accessible chromatin regions. Proceeding to a broader scale, akin to looking at multiple exons that make up a gene, we explore for mutations across the cistromes of transcription factors in accessible chromatin regions of luminal breast cancer. Together, our study identified cancer driver cistromes assigned to transcription factors essential to luminal breast cancer.

Material and Methods

Patient tumor samples

Twenty-six primary tumors with cancer cell content higher than 60% estimated by a central pathologist using hematoxylin and eosin staining were obtained from surgical specimens of patients with ER+PR+ invasive ductal carcinoma. Patients’ consent and tumor stratification were obtained through University Health Network (UHN) living biobank under REB # 16–5524.

Tumor processing and Assay for Transposase-Accessible Chromatin using sequencing library preparation

Breast tumors were minced into small pieces and digested at 37°C in mammary Epicult (STEMCELL Technologies) supplemented with 10% FBS (WISENT) and collagenase (STEMCELL Technologies), and further dissociated in 5 mg/mL dispase for 2 minutes. Cells were counted and live cells sorted into two populations, immune and epithelial, using a BD FACS aria (BD Biosciences), which were lysed for 5 minutes followed by transposase reaction and library amplification using Nextera DNA Library Prep Kit (Illumina). Libraries were then size-selected (240–360 bp) using PippinHT (Sage Science) and sequenced (NextSeq 500) using 50 bp single reads at a coverage of 40 to 80 million reads.

ATAC-seq and data analysis

Reads were aligned to hg19 using bowtie2/2.0.5 using default parameters. Aligned reads were then filtered by removing duplicated and mitochondrial reads using samtools/0.1.18. We then used MACS2/2.0.10 (30) to call accessible chromatin peaks using the following parameters: macs2 callpeak -t (input.bam) -g hs -keep-dup all -n (sample-name) -B -nomodel -SPMR -q 0.005 –outdir {OutputDir}. Peaks from all of the samples were then merged to create a catalog of accessible chromatin using bedtools merge using the following parameters: cat "_peaks.narrowPeak | bedtools sort -i stdin | bedtools merge -i stdin > Catalogue.bed.

We then generated a signal matrix by mapping the max peak signal from each sample to the full catalog of accessible chromatin using the map tool from bedtools v2.27.1.

Enrichment of genomic features in accessible chromatin regions

The accessible chromatin regions from ATAC-seq, represented using a BED file, were used as input for cis-regulatory element annotation system (CEAS) v1.0.2 (31) along with hg19 refGene, running the default chromatin immunoprecipitation (ChIP) Region Annotation and Gene-centered Annotation modules. Similarity between ATAC-profiles was estimated using all unique peaks in a pair-wise comparison between samples. A cosine similarity was used to negate the differences in global peak amplitudes and compare the relative amplitudes.

Motif enrichment

We analyzed motif enrichment using CentriMo from the Meme-suite tool version 4.9.0.4 and as a reference, we used the JASPAR_CORE_2016.meme database. This analysis was run on multiple catalogs. First, we run PM-Lum and TCGA_Lum catalogs using as a background a catalogue of publicly available DNAseq sensitive sites identified in several cell lines. The DNAseq sensitive sites were downloaded from the Encyclopedia of DNA Elements (ENCODE). Next, we ran the same analysis on PM_Lum accessible chromatin regions that overlapped mutations from International Cancer Genome Consortium (ICGC), EU and ICGC_US datasets using as a background the full PM_Lum Catalogue.

Hotspots of cis-regulatory, significantly-mutated elements

In order to identify mutation enrichment within noncoding regions, we developed an algorithm that uses an exact binomial test for each region of interest against a sample-wide noncoding background mutation rate (https://github.com/puglab/BCa_ATACSEQ_Project/tree/main/HoRSE). We first define the search space as the overlap between cis-regulatory elements and the ATAC catalog, as well as separate variants into noncoding and coding based on the University of California Santa Cruz (UCSC) hg19 known gene annotations. By tiling a 5 kb window across the cis-regulatory elements for the search space, we fit the number of variants found within the tiled cis-regulatory elements to a poisson model to estimate the average background mutation rate. We also used a 5 kb sliding window approach to identify the loci within each cis-regulatory element with the highest mutation burden. The highest mutation burdens were compared with the background mutation rate using an exact binomial test and corrected for multiple hypothesis testing using an FDR correction.

Mutation enrichment at motif sites

To analyze the enrichment of mutations at motif sites, we used a modified version of the previously published tool mutation enrichment at motif sites (modMEMOS; https://github.com/puglab/BCa_ATACSEQ_Project/tree/main/modMEMOS; ref 7; Supplementary Fig S3). First, similar to the previous version, we scanned for motif sites using either the PM_Lum ATAC-seq Catalog or PM_Lum ATAC-seq Catalog that overlap publicly available ChIP sequencing.
(ChIP-seq) data run on MCF7 using MOODS/1.9.2 tool (32). The previously published version of MEMOS assumed a normal distribution of number of mutations, however, due to the low number of mutations within cis-regulatory elements, we adopted a poisson distribution to better fit our data. Additionally, MEMOS established the null distribution of mutation enrichment by randomly sampling from the entire genome followed by adding a flanking region, resulting in the potential for the background region to include the target regions. We address this by adding the maximum flanks (1,000 bp) to all motif recognition sites first, and then restricting sampling to all regions that do not overlap the ENCODE blacklist regions as well as all original motifs ± 1,000 bps. Finally, MEMOS estimates its $P$ value for motif enrichment by calculating the distance of the number of mutations within the target cistromes from the standardized mean of the null distribution. Due to the low number of mutations within some of our cistromes, we opted for a confidence interval approach by resampling the target and background regions, followed by calculating mutation enrichment within the resampled regions and estimating the effect size of enrichment using Cohen $D$.

From a technical perspective of modMEMOS, we added a flanking region (0–1,000 bp) to Motif sites/ChIP peak centers using Bedtools slop and resampled the resulting bedfiles 100 times, taking 80% of the bedfiles each time. In parallel, we generated a background bedfile by randomly shuffling all of the motif sites ±1,000 bp while excluding the Motif sites/peak center ± flanking region as well as the ENCODE blacklist regions. Similar to Motif sites/ChIP peak centers, the background bedfile was resampled 100 times, taking 80% of the regions each time. Taking into consideration our regions of interest and background file we identified the regions that overlapped mutations from ICGC-EU and US datasets, and counted the number of mutations for each transcription factor site and flanking region. Finally, we compared the mutation counts from the region of interest with the background and calculated Cohen $D$ using the following equation: “Mean difference/pooled SD". We determined the enrichment threshold based on the Cohen $D$ median.

Intra-genomic replicates

To predict the effect of single-nucleotide variant (SNV) on transcription factors binding affinity, we run the Intragenomic replicates (IGR) tool (15). In summary, IGR uses ChIP-seq data of the transcription factor of interest to analyze the change in signal intensity in regions harboring SNVs compared with surrounding regions. Herein, we analyzed the binding affinity of the transcription factor that binds sites found to be enriched in mutation. Our regions of interest were the transcription factor binding sites ± 100 bp flanking regions. We used the ICGC-EU mutation dataset as the SNV file.

Essentiality screens

Project Achilles genome-wide short hairpin RNA (shRNA) essentiality screen data was downloaded from the DepMap portal, specifically the “Achilles” dataset (33, 34). The analysis was focused on breast cancer cell lines that showed consistency in subtyping according to all three genestsets PAM50, SCMD2, and SCMGNE. The probability of essentiality was used as a score 1 being most essential and 0 nonessential.

Identifying luminal-specific essentiality

Enrichment of essentiality for one breast cancer type compared with the rest was calculated using an approach inspired by gene set enrichment analysis (GSEA; ref. 35). The probability of essentiality ($P_e$) values were assigned a direction based on whether they were part of the cancer type of interest (COI; positive) or not (negative). A curve was fitted to the ordered $P_e$ list and the AUC was calculated. An exact $p$ value for each cancer type was calculated using a permutation test ($n_{perm} = 1,000$) where the cancer type index was randomized and the AUCs recalculated. All $p$ values were corrected for multiple testing using FDR. The standardized AUC was calculated based on a minimum (min)/maximum (max) AUC range, where the min is defined as $P_\text{min} = -1$ for all non-COIs and $P_\text{max} = 0$ for all COIs, while the max has $P_\text{max} = 0$ for all non-COIs and $P_\text{max} = 1$ for all COIs.

Ethics approval and consent to participate

The UHN Ethics Board operates in compliance with the Tri-Council Policy Statement reviewed and approved this project REB #16–5524.

Availability of data and material

The ATAC-seq raw data generated from our PM_Lum cohorts were uploaded to European Genome-phenome Archive (EGA) under accession code: EGAS00001005235. Availability of codes: https://github.com/pugliah/BCa_ATACSEQ_Project

Results

Comprehensive chromatin accessibility analysis in primary ER$^+$ PR$^+$ luminal breast cancer

To identify cis-regulatory elements, we used ATAC-seq (29, 36) to map the accessible chromatin of 26 luminal primary ER$^+$ PR$^+$ invasive ductal carcinomas breast tumors freshly collected at the Princess Margaret Cancer Centre (PM_Lum; $n = 26$; Supplementary Table S1). To enrich for malignant cells, we used flow cytometry to sort cells from dissociated tumors using the anti-CD45RO (anti-CD45) antibody (Fig. 1A). In the immune-depleted (CD45$^-$) cancer cells, we identified a catalogue of 99,516 (41.37 Mb) unique nonduplicated regions found in accessible chromatin as defined by ATAC-seq peak coverage called using MACS2 (ref. 30; Supplementary Table S2). Furthermore, we observed that 98% of peaks from our catalog were found in more than half of the PM_Lum cohort (Supplementary Fig. S1A). Moreover, we examined the size distribution of regions of accessible chromatin in the PM_Lum samples compared with our catalog, we show that the size distribution seen in catalog is similar to the one seen in the individual samples (Supplementary Fig. S1B, top). This similarity suggests that our catalog is recapitulating an accurate ATAC-profile to those seen in our samples, rather than being skewed to smaller peaks. To profile this similarity further, we calculated the fraction of overlap between our catalogue and PM_Lum cohort and show an average of 58.7% ± 12.4% overlap (Supplementary Fig. S1B, bottom).

To examine the quality of our data, we ran a similarity pair-wise comparison between accessible chromatin profiles using cosine similarity metric. Our data indicated a high degree of agreement of accessible chromatin distributions between our PM_Lum samples (Cosine similarity $\mu_{cos} = 0.82 ± 0.07$; Cosine similarity; Fig. 1B). To identify whether our catalog of accessible chromatin regions was representative of other ER$^+$ PR$^+$ breast tumors, we leveraged TCGA ATAC-seq data derived from nonenriched bulk ER$^+$ tumor tissues ($n = 41$; TCGA_Lum; ref. 37). Compared with our cohort, TCGA_Lum showed a higher number of unique accessible chromatin regions (272,291 peaks; 289.89Mb) that encompassed 93.6% (93,172/99,516 peaks) of our PM_Lum catalogue. Of note, the PM_Lum accessible chromatin regions represented only 34.2% of the TCGA_Lum catalog (Fig. 1C), suggesting our depletion of immune...
Figure 1.
Identifying chromatin accessibility in ER\(^\text{+}\)PR\(^\text{+}\) breast cancer. A, Primary tumors were minced and dissociated for subsequent flow sorting into immune and epithelial cell populations, followed by ATAC-seq profiling. B, Heatmap showing similarities between ER\(^\text{+}\)PR\(^\text{+}\) accessible chromatin profiles. Cosine similarity analysis was calculated using comparing all chromatin accessibility of samples with each other. Bar plot showing number of called peaks per sample. C, A graph showing the chromatin accessibility saturation curve. A nonlinear regression model analysis was performed using the number of unique ATAC peaks discovered in each sample to estimate the percentage of accessible chromatin mapped in 1,000. D, A bar plot showing the number of accessible chromatin regions from TCGA_Lum datasets that overlapped PM_Lum in blue and the ones unique to each cohort in red. E, A graph showing the percentage of distribution of mapped accessible chromatin regions within the genome. The cis-regulatory element annotation system (CEAS) is utilized to perform genomic distribution analysis of the accessible chromatin region mapped by ATAC-seq. F, A graph showing the relative similarity of TCGA_Lum to PM_Lum. G, A bar plot showing the number of target peaks for each motif.
cells may have enhanced the signal specific to cancer cells. Consistent with this observation, we estimated that our analysis led to the mapping of 88% of accessible chromatin within our cohort of 26 samples while the TCGA_Lum cohort reached similar saturation (87%) with 41 samples (Fig. 1D). Thus, we established a catalog from our PM_Lum cohort of high-confidence accessible chromatin regions that were found across our cohort and illustrate a high level of robustness by being found almost entirely within the independent TCGA_Lum catalog.

We next characterized the genomic distribution of accessible chromatin regions across different genomic features (e.g., promoters, distal regions, coding exons untranslated regions (UTR), and intronic regions) within the PM_Lum and TCGA_Lum catalogs. Using the CEAS tool to estimate the relative enrichment level of accessible chromatin in regions in gene features (31), we found that on average 36% of cis-regulatory elements mapped to promoters, 23% to introns, 23% to intergenic regions, 14% to UTR, and 3% coding exons (Fig. 1E). Using this same approach, we found that the accessible chromatin regions captured in the ATAC-seq data from the TCGA_Lum cohort had a similar distribution of intergenic regions (PM_Lum = 23%, TCGA_Lum = 26%; \( P = 0.10, \) two-sided \( t \) test) and coding exons (PM_Lum = 3%, TCGA_Lum = 3%; \( P = 0.42, \) two-sided \( t \) test). However, in contrast to the PM_Lum cohort the TCGA_Lum shows a higher distribution to introns (TCGA_Lum = 39%, PM_Lum = 23%; \( P < 0.001, \) two-sided \( t \) test) and a lower distribution to promoters (TCGA_Lum = 26%, PM_Lum = 36%; \( P < 0.001, \) two-sided \( t \) test) and UTRs (TCGA_Lum = 6%, PM_Lum = 14%; \( P < 0.001, \) two-sided \( t \) test; Fig. 1E). To assess the concordance of these patterns with two widely used cell lines, we examined the genomic distribution of accessible chromatin using DNase I hypersensitive sites from both MCF7 and T47D cells. This analysis found, relative to both cohorts of primary tumours, the cell lines had a higher percentage of open chromatin regions in distal intergenic regions and lower percentage in promoter regions (Fig. 1E). Since the accessible chromatin regions differed significantly from our set of primary tumours, we did not include them in subsequent analyses. Thus, our results highlight that both PM_Lum and TCGA_Lum accessible chromatin catalogs favor noncoding regions, where most accessible regions are found in the promoter, intergenic, and intronic sequences as opposed to the coding exons.

Considering that we used cell sorting to exclude immune cells from our tumor samples using an anti-CD45 antibody, we examined whether the difference in accessible chromatin profiles that we saw between TCGA_Lum and PM_Lum was due to immune infiltration. We tested for this immune infiltrate by comparing the similarity of the PM_Lum and TCGA_Lum profiles to a known immune reference comprised of publicly available chromatin accessibility data (DNaseI) from 12 immune-cell types [trophoblast, CD1c\(^{-}\), myeloid progenitors, CD14\(^{-}\) monocytes, T helper1, T helper1, T helper2, CD8\(^{+}\) T cells, naive thymocytes T cells, CD4\(^{+}\) \( \alpha \)-\( \beta \) T cells, natural killer (NK) cells, and B cells]. Our results showed that the TCGA_Lum chromatin accessibility profile was significantly more similar to the accessible chromatin profile for 9 of the 12 immune-cell types (trophoblast, CD1c\(^{-}\), myeloid progenitors, CD14\(^{-}\) monocytes, T helper1, T helper1, T helper2, CD8\(^{+}\) \( \alpha \)-\( \beta \) T cells, naive thymocytes T cells) compared with the PM_Lum profile (Fig. 1F; \( P < 0.001, \) one-sided \( t \) test). The accessible chromatin profile for 3 of the 12 immune cells tested (CD4\(^{+}\) \( \alpha \)-\( \beta \) T cells, NK, and B cells) were not significantly different from the fact that CD45RO is not expressed in CD4\(^{+}\) T cells, NK, and B cells (38). Altogether, our data suggests that although there are similarities between TCGA_Lum and PM_Lum, the cell sorting performed on our PM_Lum cohort led to a depletion of immune cells, resulting in a more cancer cell–enriched accessible chromatin catalog.

Cis-regulatory elements work through the recruitment of transcription factors that bind to unique DNA recognition sequences. We therefore assessed the sequence composition of accessible chromatin regions from ER\(^{+}\)PR\(^{+}\) breast tumors through DNA recognition motif enrichment analysis. Using the JASPAR database as a reference for motif recognition sites and the pan-cancer ENCODE DNase I hypersensitive sites as a background, we utilized the CentriMo method to identify 40 significantly enriched DNA recognition motif families, 6 of which are known to play an important role in luminal breast cancers: AP-2 (TFAP2A), Forkhead (FOXA1), STAT (STAT3), C/EBP (CREBBP), NR1 (RORA), GATA (GATA3; refs. 15, 39–41; \( P < 0.001; \) Fisher exact test; Fig. 1G; Supplementary Table S3.1). To corroborate our findings, we performed a similar DNA recognition motif enrichment analysis on the TCGA_Lum catalog. We identified 57 DNA recognition motif families enriched in this cohort; 33 of 57 overlapped with the motifs enriched in our PM_Lum catalogue, 24 of 57 were unique to the TCGA catalogue, and 7 of 40 (HSF, MyoD/DASC, RH1, MAD5, NEAT, NF, and B-ATF) were found only in the PM_Lum catalogue (Supplementary Fig. S1C; Supplementary Table S3.2). All of which have been linked to breast cancer development, tumor invasion, and drug resistance (39, 42–45). Together, these results demonstrated that our PM_Lum catalog defines a broad spectrum of motif recognition sites, 82.5% of which are also found in the TCGA_Lum catalogue and 6 which are established markers of luminal breast cancer biology, thus reflecting the luminal breast cancer specificity of our catalog.

**Individual cis-regulatory elements are rarely recurrently mutated**

The enrichment of mutations within promoters and enhancers of key breast cancer genes, such as TERT (6, 24) and FOXA1 (18), suggests potential for recurrent mutations in additional regulatory regions. To search for other mutations in cis-regulatory elements in ER\(^{+}\)PR\(^{+}\) breast cancer, we integrated our PM_Lum catalogue with somatic mutations from 348 ER\(^{+}\)PR\(^{+}\) breast cancers in two WGS breast studies (ICGC-EU; ref. 4; \( n = 306 \) and ICGC-US; ref. 46; \( n = 42 \)). Of the 1,048,537 mutations found across WGS of ER\(^{+}\)PR\(^{+}\) breast cancer samples from ICGC-EU and ICGC-US, an average of 1.7% (ICGC-US = 1.76%; ICGC-EU = 1.78%; 0.7%–3.4%; \( n_{SNVs} \) min = 4,295, max = 35,650) were detected within our PM_Lum catalog, which comprises 1.3% of the genome (Fig. 2A). To identify whether our PM_Lum catalog captured mutations specific to ER\(^{+}\)PR\(^{+}\) breast cancers, we compared the localization of mutations to 19 ICGC WGS cancer cohorts (Supplementary Table S4). We found that these additional 19 cancer types all had significantly lower fractions of mutations overlapping our PM_Lum catalog as compared with ER\(^{+}\)PR\(^{+}\) breast cancer samples, with the exception of BOCA, PAEN-AU, and PRAD-UK (Fig. 2A). We then performed the same analysis using the TCGA_Lum catalog of accessible chromatin and found similar results, with luminal breast tissue having a higher percentage of mutations localized to this region when compared with other tissues (\( P < 0.01, \) two-sided \( t \) test; Supplementary Fig. S2A).

These results highlight that mutations with luminal breast cancers are predominantly found within our accessible chromatin catalog, thus setting the stage for interpreting mutations in cis-regulatory elements relevant to luminal breast cancer biology.

To identify highly mutated regulatory elements in ER\(^{+}\)PR\(^{+}\) breast cancer, we analyzed frequently mutated regulatory elements using the ActiveDriverWGS method (10). Restricting our analysis to our
PM_Lum catalog as the target regions, we found no driver mutations after multiple testing correction using two separate datasets, ICGC-EU (Supplementary Fig. S2B, Supplementary Table S5.1) and ICGC-US (Supplementary Fig. S2C, Supplementary Table S5.2) WGS data \( (q < 0.01; \text{FDR}) \). By running a similar analysis on the TCGA_Lum catalog, ActiveDriverWGS identified one highly mutated distal region \( (\text{chr10:8115662–8116163}) \) using ICGC-EU (Supplementary Fig. S2D) and none using ICGC-US (Supplementary Fig. S2E). Although ActiveDriverWGS is a robust tool for calling drivers in regulatory elements, it takes a conservative one-to-one approach between mutations and active elements, negating the cumulative effect of multiple mutations within a hotspot region. To address these limitations, we designed an algorithm Hotspot of cis-Regulatory, Significantly-mutated Elements (HoRSE) that relaxes the stringency of ActiveDriverWGS by looking for clusters of hotspot mutations within regulatory elements against a background of global and local somatic mutation rates (Supplementary Fig. S2F; Online methods). Using HoRSE, we found 5 unique cis-regulatory elements enriched for somatic mutations across ICGC-EU and -US \( (n_{\text{ICGC-EU}} = 5, n_{\text{ICGC-US}} = 1; q < 0.01, \text{exact binomial test}) \) with PLEKHS1 being the only cis-regulatory element significantly enriched in both WGS cohorts (ICGC-EU: \( n = 12/308 \); ICGC-US: \( n = 6/42 \); Fig. 2B and C; Supplementary Table S5.4). Two of the somatic mutations identified within the PLEKHS1 promoter are thought to be attributed to APOBEC DNA-editing activity \((4, 47)\). In the ICGC-EU dataset, we identified 4 cis-regulatory elements enriched for somatic mutation in addition to PLEKHS1 (Promoters of INTS2; \( n = 6/308 \), APLP1; \( n = 6/308 \), and CCDC107/RMRP; ref. 18, 47; \( n = 6/308 \), and Distal Region: chr1:129512774–129513782; \( n = 7/308 \); Fig. 2B and C; Supplementary Table S5.3). We calculated the number of samples within our cohort with ATAC-seq coverage of these mutationally enriched regions and show that they are within accessible chromatin seen across several samples (PLEKHS1; \( n = 4/26 \); INTS2, \( n = 15/26 \); APLP1, \( n = 15/26 \); CCDC107/RMRP, \( n = 25/26 \); chr1:129512774–129513782; \( n = 15/26 \); Supplementary Table S6). Additionally, by applying our algorithm on the regions covered by the TCGA_Lum catalog, we revealed 19 significantly mutated regions in the ICGC-EU dataset regions including CCDC107/RMRP (Supplementary Fig. S2G) and 3 regions in ICGC-US (promoter: RARA and 2 distal regions: chr8:98131092–98131993 and chr17:38603438–3860433; Supplementary Fig. S2H). Our results highlight the small number of recurrent mutational hotspots across all the cis-regulatory elements of luminal breast cancer. Thus, similar to how the search for driver genes is hindered when focusing on single exons, our results show the hunt for cancer drivers within individual cis-regulatory elements may be too limiting resulting in the few observed recurrently mutated regions.

Noncoding mutations reveal cancer driver cistromes in luminal breast cancer

The genome can be looked at as a collection of cis-regulatory elements that can be organized into cistromes, based either on the DNA recognition sequence content or on actual occupancy by transcription factors. As our previous analysis highlights the limitations of identifying drivers using individual cis-regulatory elements, our next
step was to assess the presence of cancer driver cistromes in ER⁺ PR⁺ luminal breast tumors. First, we measured the enrichment for DNA recognition motifs within the PM_Lum catalog of cis-regulatory elements found to be mutated in primary luminal breast tumors from the ICGC-EU and -US studies. This revealed significant enrichment for several DNA recognition motifs related to the JUN, FOS, Forkhead, NFAT, POU, and REL families of transcription factors across both ICGC dataset (Fig. 3A). The NF1, C2H2, IRF, and HD-CUT DNA recognition motifs were uniquely enriched in cis-regulatory elements mutated based on the ICGC-EU dataset (Fig. 3A).

To focus on DNA recognition motif-based cistromes relevant to luminal breast cancer, we subdivided cis-regulatory elements from our catalogue of accessible chromatin regions based on the presence of DNA recognition motifs enriched in mutated cis-regulatory elements across both ICGC-EU and ICGC-US datasets, namely JUN, FOS, Forkhead, NFAT, POU, or REL. We calculated the frequency of mutations across varying window sizes (0 to 1,000 bp) around the cis-regulatory elements from each of the motif-based cistromes using modMEMOS (modMEMOS and Flanking Regions; Supplementary Fig. S3; Online methods; refs. 7, 18). We estimate the effect size of mutation enrichment in DNA recognition motifs compared with a background model using Cohen D, a statistical value that represents the standardized difference between two means. Using a window of 50 bp flanking the motif recognition sites, as defined by the work from Mazrooei and colleagues (7, 18), we found an enrichment for mutations near the JUN, FOS, and Forkhead motif-based cistromes in both ICGC-EU (Fig. 3B) and ICGC-US data sets (Fig. 3C). Additionally, cis-regulatory elements proximal to POU motif cistrome were found to be enriched in mutations uniquely in the ICGC-EU dataset (Fig. 3B). These results suggest that noncoding mutations preferentially accumulate across cis-regulatory elements that harbor specific DNA recognition motifs, namely JUN, FOS, or Forkhead motifs.

Given that transcription factors of the same family can bind the same DNA recognition motif, we explored the transcription factor–

![Figure 3](image-url)

**Figure 3.** Mutation analysis at recognition sites of motifs enriched in ER⁺ PR⁺ breast cancer. **A,** Lollipop graph showing enriched motif families in PM_Lum catalog overlapping SNVs from ICGC-EU (red) and ICGC-US (blue) against the total PM_Lum catalog (p value < 0.01; grey; p value > 0.01, Fisher exact test). **B** and **C,** graph (top) and heatmaps (bottom) showing the enrichment of mutations at DNA recognition sites found to be significantly enriched in the PM_Lum catalog using ICGC-EU (**B**) and ICGC-US (**C**) mutation calls. Cohen D was calculated based on resampling and the value indicates significant enrichment. The red dotted line indicates Cohen D median.
based cistromes to examine whether these variants are targeting transcription factor–binding sites specific to breast tumors. We leveraged the publically available collection of ChIP-seq datasets of transcription factors \((n = 48)\) and cofactors \((n = 30)\) from the luminal breast cancer cell line \(\text{MCF7; ref. 48}\) to identify luminal specific transcription factor–based cistromes. We first clustered all cistromes according to their similarity in ChIP-seq signal across our catalog of cis-regulatory elements from luminal breast tumors and identified 7 distinct clusters (Supplementary Fig. S4), including one consisting of the ER, FOXA1, and GATA3 transcription factors \(\text{TFs}_1\). We next used modMEMOS to quantify the enrichment of mutations over these cistromes. Using the mutation calls from the ICGC-EU dataset, we identified 28 cancer driver cistromes \(\text{AHR, AR, CEBPB, CREBBP, CTCF, ELF1, ER, FOSL2, FOXA1, FOXM1, GABPA, GATA3, JUND, MAX, MYC, NR2F2, REST, TCF12, TEAD4, TFAP2A, TFAP2C, and ZNF217; Fig. 4A}\). We further refine these transcription factor–based cistromes by including only the cis-regulatory elements that harbor a matched DNA recognition site for the designated transcription factor family. Using modMEMOS on these DNA recognition site–specific transcription factor–based cistromes, we identified 10 cancer driver cistromes \(\text{CTCF, ELF1, ER, FOSL2, FOXA1, FOXM1, GATA3, JUND, TFAP2A, and TFAP2C}\) that are enriched in mutations both in the ICGC-EU \(\text{Fig. 4B}\) and ICGC-US \(\text{Fig. 4C}\) datasets. Consistent with the motif-based cistromes that we identified as cancer drivers \(\text{Fig. 3B and C}\), we observed a similar enrichment of motif, but not all transcription factor–based cistromes that compose each motif family \(\text{Forkhead, JUN, and FOS}\), with the exception of the REL motif family (Supplementary Fig. S5). Furthermore, we found that in the majority of cases, not all transcription factor–based cistromes for a given motif family defined cancer driver cistromes \(\text{e.g., Forkhead motif family}\). Rather mutations were found to be enriched in specific transcription factor–based cistromes \(\text{Supplementary Fig. S5}\). Altogether, our results highlight that key transcription factor–based cistromes are cancer drivers independent of their motif families or based on their similarity to other cistromes, indicating the mutations are selectively enriched within specific driver cistromes.

We next examined if the noncoding mutations within or flanking \((100 \text{ bp})\) the DNA recognition motif found within the cancer driver cistrome for \(\text{CTCF, TFAP2C, GATA3, FOXA1, ER, FOSL2, JUND, TFAP2A, ELF1, and FOXM1}\) could alter transcription factor binding to the chromatin. Using the IGR method \((15)\) predicted that less than 40% of the noncoding mutations could alter the binding intensity of any of these transcription factors to the chromatin \(\text{CTCF: Down = 36\%, Up = 15\%; TFAP2C: 31\%,28\%; GATA3: 19\%,10\%; FOXA1: 14\%,17\%; ER: 18\%,9\%; FOSL2: 27\%,13\%; and JUND: 14\%,5\%; TFAP2A: 32\%,20\%; ELF1: 33\%,18\%; FOXM1:20\%,13\%; Supplementary Fig. S6}\). These results argue that despite the enrichment of mutations observed over transcription factor–based cistromes, only a minority of these mutations can directly impact the binding affinity of transcription factors to cis-regulatory elements.

**Cancer driver cistromes correspond to transcription factors essential to luminal breast cancer**

To better understand why specific transcription factor–based cistromes are enriched for noncoding mutations in luminal breast cancer, we examined whether this enrichment reflected the dependency to some as opposed to all transcription factors expressed in luminal breast cancer. Using the genome-wide shRNA essentiality screen data from luminal breast cancer cell lines generated as part of the DepMap project \((33, 34)\), we found that 4 of the 10 transcription factors linked to cancer driver transcription factor cistromes were exclusively essential in luminal breast cancers \(\text{GATA3, ESRI, FOXA1, TFAP2A}\) and 5 additional transcription factors were essential in all breast cancers, regardless of subtype \(\text{CTCF, FOXM1, TFAP2C, JUND, and FOSL2; Fig. 5, P < 0.05}\). ELF1 was the only transcription factor linked to a cancer driver cistrome not essential in luminal breast cancer cells \(\text{Fig. 5}\). While we found that the CREBBP and CEBPG transcription factors were essential preferentially in luminal breast cancer, we did not identify these transcription factor cistromes as cancer drivers as they were only significantly enriched in mutations in the ICGC-EU dataset. Altogether these results support the identification of cancer driver cistromes based on transcription factors that are essential to the growth of luminal breast cancer.

**Discussion**

Our study depicts the cancer driver cistromes specific to luminal breast cancers as identified by an enrichment of noncoding mutations flanking DNA recognition motifs of cis-regulatory elements accessible in luminal breast tumors. Using flow-sorting to enrich the cancer cell population, we generated a robust catalogue of luminal-enriched accessible chromatin regions. Within this catalogue, we identified seven recurrently mutated cis-regulatory elements that occur at a low frequency. By expanding our search to transcription factor–based cistromes, we identified 10 cancer drivers and showed that a minority of the noncoding mutations can directly impact the transcription factor binding to cis-regulatory elements. Finally, we show that 9 out of the 10 transcription factor cistromes are essential to breast cancer, and 4 of which are specific to luminal breast cancer.

Somatic variants and genomic rearrangements affecting the protein-coding regions of luminal breast cancers have been well characterized \((4, 5, 49, 50)\), these regions account for less than 2% of the genome \((51, 52)\). The importance of acquired genetic variants found in cis-regulatory elements is highlighted in a luminal breast cancer study by Bailey and colleagues \((8)\) and across multiple breast cancer subtypes by Rheinbay and colleagues \((18)\). Bailey and colleagues identified several somatic mutations with functional consequences within the promoters and enhancers that regulate the \(\text{ESRI}\) gene \((8)\). The study by Rheinbay and colleagues describes somatic mutations across several promoters, including \(\text{FOXA1}\), and their effect on gene expression \((18)\). Our analysis of the mutation burden within luminal \(\text{ER}^+\) breast cancer cis-regulatory elements yielded only seven significant hits. Across both the ICGC-US and -EU cohorts, we found significant enrichment of mutations in the \text{PLEKHS1}\ promoter that is likely a result of APOBEC DNA-editing activity \((4)\), however, mutation within this region lead to an increase in \text{PLEKHS1}\ gene expression \((28)\) and it’s also known as a genetic marker of aggressiveness for differentiated thyroid carcinomas \((53)\). Although significant, our results show that the hunt for cancer drivers within individual cis-regulatory elements is limiting at best, resulting in the few observed recurrently mutated individual cis-regulatory elements. Discovering cancer driver mutations in the noncoding space is challenging due to heterogeneity in the cis-regulatory element and mutational space between individual tumors, leading to a need of large datasets to identify rarely occurring cancer driver mutations \((52)\).

As individual cis-regulatory elements are functional units of the cistrome, akin to how exons make up a gene, we expanded our search for cancer drivers by partitioning our accessible chromatin region into cistromes specific for luminal breast cancer. Genome-Wide Association Studies \((\text{GWAS})\) have identified thousands of risk variants linked to diseases including breast cancers \((7, 8, 15, 18, 54)\). In luminal breast cancer a number of these risk variants have been shown to accumulate
at the cistromes of key transcription factors in luminal breast cancer, namely ER and FOXA1 (8, 15, 55). The CTCF/cohesin binding sites, regulators of the 3D structure of chromatin, are enriched in point mutations in a highly stereotypic pattern across various cancer types which may affect transcriptional regulation and result in genomic instability (56). Additionally, Mazrooei and colleagues showed enrichment of mutations within the cistrome of master regulators of prostate cancer such as FOXA1, HOXB13, and AR (7). Our study provides a look into another aspect of cancer-driver search by looking at mutation load within motif and transcription factor–based cistromes. We detected an enrichment of mutations within the cistrome of master regulators of prostate cancer such as FOXA1, HOXB13, and AR (7). Our study provides a look into another aspect of cancer-driver search by looking at mutation load within motif and transcription factor–based cistromes. We detected an enrichment of mutations within the cistrome of master regulators of prostate cancer such as FOXA1, HOXB13, and AR (7). Our study provides a look into another aspect of cancer-driver search by looking at mutation load within motif and transcription factor–based cistromes. We detected an enrichment of mutations within the cistrome of master regulators of prostate cancer such as FOXA1, HOXB13, and AR (7). Our study provides a look into another aspect of cancer-driver search by looking at mutation load within motif and transcription factor–based cistromes. We detected an enrichment of mutations within the cistrome of master regulators of prostate cancer such as FOXA1, HOXB13, and AR (7). Our study provides a look into another aspect of cancer-driver search by looking at mutation load within motif and transcription factor–based cistromes. We detected an enrichment of mutations within the cistrome of master regulators of prostate cancer such as FOXA1, HOXB13, and AR (7). Our study provides a look into another aspect of cancer-driver search by looking at mutation load within motif and transcription factor–based cistromes. We detected an enrichment of mutations within the cistrome of master regulators of prostate cancer such as FOXA1, HOXB13, and AR (7). Our study provides a look into another aspect of cancer-driver search by looking at mutation load within motif and transcription factor–based cistromes.
the cistromes of CTCF, TFAP2C, GATA3, FOXA1, ER, FOSL2, JUND, TFAP2A, ELF1, and FOXM1. The biological significance of mutagenic processes occurring at the flanking regions of cistrome over the active binding sites is yet to be fully understood but is a phenomenon seen in prostate cancer (7). While other studies in melanoma (57, 58), lung (59), and colorectal (56) cancers have found the inverse true, they have attributed this mutational enrichment to restricted DNA-accessibility affecting repair machinery due to either chromatin conformation change, or occupancy of specific transcription binding sites by proteins (60). Approximately 5% to 36% of these mutations are predicted to impact transcription factor binding to the chromatin. Altogether, we describe an increase of mutational burden at specific cistromes defining them as cancer driver cistromes.

As validation of our cancer driver cistromes, we determined from the DepMap project (33, 34) that 4 transcription factors associated with our driver cistromes were preferentially essential to luminal breast cancers: GATA3, ESR1, FOXA1, and TFAP2A. Among those, GATA3, ESR1, and FOXA1 have been widely shown to be involved in luminal breast cancer development and resistance to endocrine therapy (61), while TFAP2A is associated with the luminal breast phenotype (41). Five additional transcription factors, CTCF, FOXM1, transcription factor AP2C, JUND, and FOSL2, were essential across all breast cancer subtypes.

Figure 5.
Cancer driver cistromes are of transcription factors essential to luminal breast tumors. A heatmap showing the probability of the essentiality of the transcription factor in several breast cancer cell lines with different subtypes (Luminal, triple-negative breast cancers (TNBC), and HER2). Column annotation indicates the enrichment of mutations at binding sites ± 50 bp, and rows annotation shows cell line subtype.

Subtype
Luminal
TNBC
Her2Amp

P
1 0.8 0.6 0.4 0.2

Luminal-specific essentiality
Yes P < 0.05
No P > 0.05

Enrichment of mutations

Subtype
NRF1
MYC
CTCF
HSF1
ZNF217
CREBBP
GABPA
TFAP2C
GATA3
FOXA1
ESR1
FOSL2
HFX3
JUND
MAPK
MNT
TFAP2A
MAZ
AR
NR2F2
NR3C1
REST
NR5A2
ELF1
CEBPG
ERG
TCF7L2
CX1
NR2F1
ELK1
FOOQ2
ESR1
KLF4
RUNX1
TCF12
AR
ARNT
FOXM1
SHP
JUN
MAX
TEAD4
CEBPG
RELA
ESRRA
SREBF1
ESE2

Subtype
HCC1428
CAMA1
MCF7
ZR751
EFM19
AU565
HCC1954
SUM229PE
MDAMB436
HS578T
BT549
HCC202
HCC1187
HCC1395
MDAMB157
MDAMB468
HCC38
SUM149PT

Enrichment of mutations
Yes
No

Cancer driver cistromes are of transcription factors essential to luminal breast tumors. A heatmap showing the probability of the essentiality of the transcription factor in several breast cancer cell lines with different subtypes (Luminal, triple-negative breast cancers (TNBC), and HER2). Column annotation indicates the enrichment of mutations at binding sites ± 50 bp, and rows annotation shows cell line subtype.
cell lines. While not luminal exclusive, these transcription factors have roles in breast cancer progression, aggressiveness, cell motility, modulating cancer-cell proliferation, and response to therapy (41, 62–66). In conclusion, our study provides new insights to identifying cancer drivers beyond the protein-coding space to benefit the development of precision medicine from cancer-driver events applicable to breast and other cancer types.

Authors’ Disclosures

P. Mazrooei reports other support from Canadian Institutes of Health Research (CIHR) during the conduct of the study. D.W. Cescon reports personal fees from AstraZeneca, Dynamo Therapeutics, Agenda, Puma Biotechnology, Exact Sciences, Gilead, Novartis; personal fees and other support from GlaxoSmithKline, Merck; grants and personal fees from Pfizer; and personal fees and other support from Roche outside the submitted work; in addition, D.W. Cescon has a patent for (US62/675,228) for “Methods of treating cancers characterized by a high expression level of spindl and kinetochore associated complex subunit 3 (aka) gene” issued. M. Lupien reports grants from CIHR during the conduct of the study. T.J. Pugh reports grants from RocheGenentech; personal fees from Canadian Pension Plan Investment Board, Chrisyalis Biomedical Advisors, Illumina Inc., Merck; and personal fees from AstraZeneca outside the submitted work. No disclosures were reported by the other authors.

Authors’ Contributions

S. El Ghamrawi: Conceptualization, resources, data curation, formal analysis, visualization, methodology, writing–original draft, project administration, writing–review and editing.

R. Quevedo: Formal analysis, methodology.

J. Hawley: Assisted in bioinformatic analysis.

P. Mazrooei: Assisted in bioinformatic analysis.

Y. Hanna: Performed macrodissection, tissue-staining, DNA/RNA extraction.

I. Cirlan: Performed macrodissection, tissue-staining, DNA/RNA extraction.

H. Zhu: Assisted in bioinformatic analysis.

J.P. Bruce: Assisted in bioinformatic analysis.

L.E. Oldfield: Resources.

S.Y.C. Yang: Assisted in tissue processing.

P. Guilhamon: Assisted in bioinformatic analysis.

J. Reimand: Assisted in bioinformatic analysis.

D.W. Cescon: Resources, S.J. Done: Resources.

M. Lupien: Conceptualization, resources, supervision, funding acquisition, investigation, writing–original draft, review–writing and editing.

T.J. Pugh: Conceptualization, resources, supervision, investigation, writing–original draft, project administration, review–writing and editing.

Acknowledgments

We thank the staff of the Princess Margaret Genomics Centre (www.pmgenomics.ca, Troy Keterla, Julissa Tsao, Nick Khour, and Monika Sharma) and Bioinformatics Services (Carl Virtanen, Zhibin Lu, Jin Qun, and Natalie Stickle) for their expertise in generating the sequencing data used in this study. This research was supported by a grant from Susan G. Komen. T.J. Pugh holds the Canada Research Chair in Translational Genomics and is supported by a Senior Investigator Award from the Ontario Institute for Cancer Research and the Gattuso-Slaight Personalized Cancer Medicine Fund at the Princess Margaret Cancer Centre. Infrastructure support was provided by the Princess Margaret Cancer Foundation; Canada Foundation for Innovation, Leaders Opportunity Fund, CIF 340 #32383; and Ontario Ministry of Research and Innovation, Ontario Research Fund Small Infrastructure Program (T.J. Pugh). This work was also supported by the CIHR (Funding Reference Number 138693, 158225, and 168933 to M. Lupien) and the Princess Margaret Cancer Foundation (to M. Lupien). M. Lupien holds an Investigator Award from the Ontario Institute for Cancer Research and the Bernard and Francine Dorval Award for Excellence from the Canadian Cancer Society. S. El Ghamrawi is supported by CIHR Banting Postdoctoral Fellowship.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked ‘advertisement’ in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 17, 2021; revised July 31, 2021; accepted September 17, 2021; published first September 22, 2021.

References


Molecular Cancer Research

Mutations in Noncoding Cis-Regulatory Elements Reveal Cancer Driver Cistromes in Luminal Breast Cancer

Samah El Ghamrasni, Rene Quevedo, James Hawley, et al.

*Mol Cancer Res* Published OnlineFirst September 23, 2021.

Updated version
Access the most recent version of this article at:
doi:10.1158/1541-7786.MCR-21-0471

Supplementary Material
Access the most recent supplemental material at:
http://mcr.aacrjournals.org/content/suppl/2021/09/22/1541-7786.MCR-21-0471.DC1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link:
http://mcr.aacrjournals.org/content/early/2021/10/16/1541-7786.MCR-21-0471. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.