Super-enhancer landscapes of ovarian cancer reveal novel epigenomic subtypes and targets

Eaton ML, Johnston B, Piotrowska K, Collins C, Lopez J, Knutson S, di Tomaso E, Carulli J, Olson E

Syros Pharmaceuticals Inc, 620 Memorial Drive, Cambridge, MA, 02139, USA

Abstract

Background: There is a critical unmet need for targeted therapies in ovarian cancer, especially high-grade serous ovarian cancer (HGSOC). We used enhancer mapping combined with transcriptomics and mutations to identify novel ovarian cancer subtypes and associated targets.

Material and methods: With ChIP-seq for H3K27Ac, we profiled the enhancer landscape of 101 primary tumor samples from 7 ovarian cancer subtypes with a focus on HGSOC, 29 cell line models, 3 PDXs, and 8 non-cancerous samples of ovarian and fallopian tube tissue. We also profiled many of these samples through RNA-seq and a focused NGS-based mutational panel. We used matrix factorization methods to reveal novel sub-groups of ovarian cancer patients and predicted their associated transcriptional circuitry.

Results: Through a computational deconvolution of enhancer maps, we identified novel enhancer defined patient subtypes of ovarian cancer. While some known subtypes, such as granulosa cell, associated uniquely with their own enhancer profile, the majority of the primary tumor samples fell into 4 clusters that did not correlate with histological subtype or with known high-frequency ovarian cancer mutations. Each cluster was associated with its own unique super-enhancer (SE) signature, implying that each is driven by a unique transcriptional circuitry. There was a striking cluster-specific patterning of many known ovarian cancer related genes such as FOXM1, CD47, and MYC, and genes linked to pathways known to be dysregulated in ovarian cancer, including an SE linked to the RB pathway gene Cyclin E1. Furthermore, many additional cluster-specific SEs were discovered representing novel potential therapeutic targets. Interestingly, while we could assign ovarian cancer cell line models to these novel subtypes, many cell lines' enhancer landscapes appeared to be distinct from those of primary tumor cells.

Conclusions: Together, our results comprise the largest ovarian cancer enhancer mapping effort to date, and demonstrate how an integrated analysis of enhancers, transcriptomes, and genotypes together can yield transcriptional circuitry that reinforces the role of known pathways associated with ovarian cancer progression and treatment, can be used to select cell models that best recapitulate the enhancer landscape of primary tumors, and can be mined to identify novel targets and biomarkers. Cluster-specific SEs at MYC and CCNE1 suggest that some tumors have increased transcriptional dependency on these loci as well as the components of the corresponding transcriptional machinery. The role of one such component, CDK7, is currently being evaluated in ovarian cancer patients with SY-1365, a first-in-class selective CDK7 inhibitor in Phase 1 clinical development (NCT 03134638).

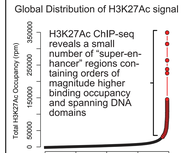
Method - Call SEs, NNMF, & cluster

Map Enhancer Regions

H3K27Ac ChIP-seq Profile at Gene Level

H3K27Ac ChIP-seq Profile Genome-wide

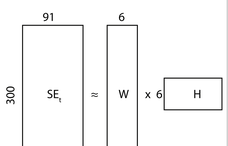
Global Distribution of H3K27Ac signal



Across 109 primary OvCa samples, select the 300 SEs with the highest across-sample variability

Identify Coordinated Patterns of Enhancer Activity

Using Non-Negative Matrix Factorization (NNMF, Brunet's algorithm), decompose transformed SE score matrix into 6 factors. The granulosa cell and yolk sack subtypes were shown to represent their own subgroups of samples, and so were held out along with the normals for de novo clustering.



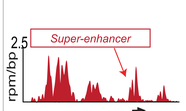
NNMF is well suited to the task since any enhancer activity in the cell population can only add to enhancer signal. The choice to use 6 factors was driven by a shoulder in the variance explained by the factors in genome-wide SEs, as well as being a reasonable number of factors given the number of samples.

Find samples with similar enhancer-pattern utilization

K-medoids clustering of the column scaled NNMF coefficient matrix (H) with euclidean distance and k=4 reveals clear clusters of enhancer pattern utilization.

Link SEs to genes

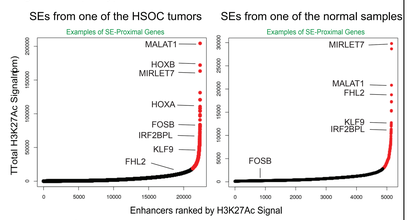
Using Pearson correlation, we link SEs to genes by searching in cis around SEs for gene expression that correlates with SE strength across samples. P-values must pass genome-wide multiple hypothesis testing correction.



Mapping H3K27Ac in primary ovarian cancer and normal tissue reveals tumor-specific Super-Enhancers (SEs)

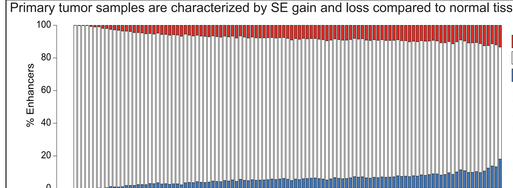
We profiled 146 SE maps across a range of sample types:

- Primary:
  - High-grade serous (HGS): 40
  - Low-grade serious (LGS): 16
  - Unknown serous (SUn): 13
  - Mucinous (Mu): 8
  - Endometrioid (En): 7
  - Yolk Sac (YS): 6
  - Granulosa cell (GC): 4
  - Unknown (Un): 4
  - Clear cell (CC): 3
- Cell lines: 29
- PDX: 3
- Normal (FT/Ov): 8
- Normal-like cell lines: 5

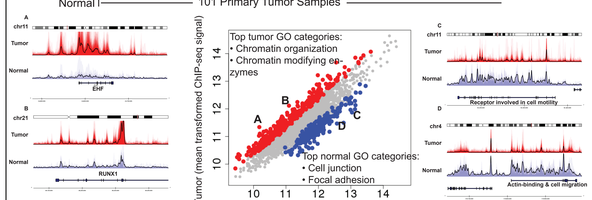


Differential enhancer activity between tumor and normal

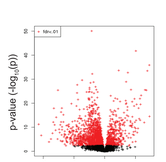
Primary tumor samples are characterized by SE gain and loss compared to normal tissue.



- Average gained SEs compared to normal: **226**
- Average lost SEs: **174**



Comparing tumor vs normal with a generalized linear model, we find 348 regions of gained SE activity, and 211 with diminished SE activity. These SEs are linked to genes that are enriched for particular pathways, the gained SEs tend to be linked to genes enriched for the GO:BP categories chromatin organization and chromatin modifying enzymes. The diminished SEs tend to be linked to genes enriched for cell junction and focal adhesion categories.

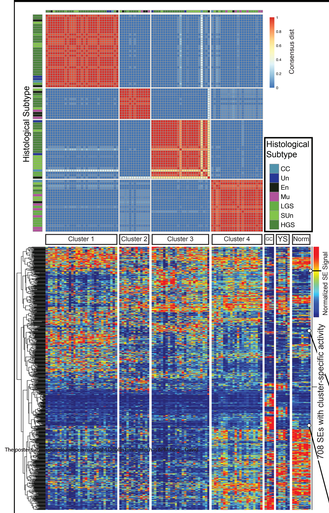


Cell lines are transcriptionally distinct from patients

Interestingly, we find that ovarian cell lines as a group are transcriptionally divergent from patients. As an example, we show an enhancer volcano plot on the left of patients vs. cell lines, revealing 1792 potential SE loci that diverge from patients with an FDR of at least 0.01 (~48% of loci).

Within the group of cell lines, we find that some cell lines are more representative than others of the general patient enhancer landscape. Importantly, we can predict a subgroup (see next panel) for each cell line.

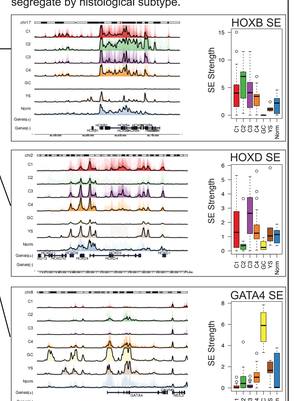
Novel ovarian cancer patient subgroup identification through NNMF



Clustering leads to subgroup-specific SE-tagged genes. The clustered samples can be queried to identify SEs that distinguish subsets of samples. The SEs (rows) are shown above, ordered by patient subgroup (columns). SEs are clustered by coordinated activity across patients (hierarchical tree, left). Three example subgroup-specific loci are shown: the HOXB locus (top), the HOXD locus (middle), and GATA4 (bottom), with lines indicating their position in the heatmap above.

Clustering patient SE maps by NNMF consensus clustering reveals 4 patient subgroups

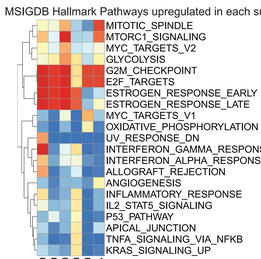
Holding out the GC and YS subtypes, we clustered the primary samples into 4 subgroups (left, top). We chose 4 as it yielded the desired complexity while maximizing the cophenetic correlation of the resulting clusters (right). These epigenome-defined subgroups did not segregate by histological subtype.

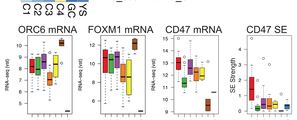


Each patient subgroup exhibits its own representative biology.

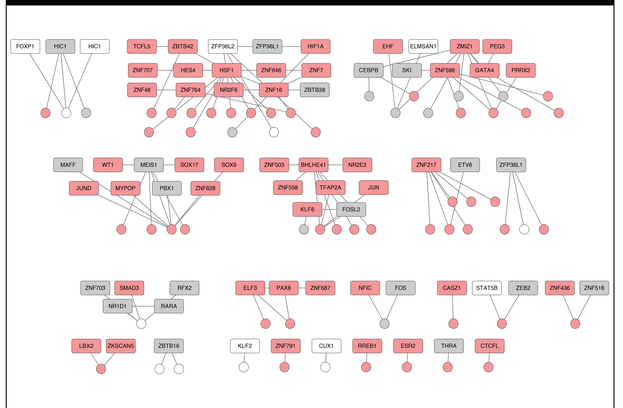
Each cluster of primary tumor samples is upregulated for unique pathways, as shown by finding GSEA enrichments for the MSIGDB Hallmark set of meta-pathways (heatmap on right, where each column shows the enrichment of the given patient cluster for a hallmark gene set). Cluster 1 exhibits higher expression of immune pathways. As an example, it has the highest expression of and largest super-enhancer linked to the CD47 transcript (bottom right), involved in immune evasion. Cluster 2 has the strongest SE at the HOXB locus, while having the weakest at HOXD, implying a distinct developmental state. Cluster 3 shows upregulated pathways involved in MTOR signaling, MYC targets, and glycolysis. Cluster 4 is characterized by increased expression in inflammatory and cytokine response pathways, but is the most "normal-like" of the clusters, with the highest SE map correlation with normal samples, as well as lower expression of FOXM1, and pre-RC components such as ORC6 (right). Interestingly, cluster membership did not correlate with known ovarian somatic mutations, implying that epigenomic subtype is independent of mutational burden.

MSIGDB Hallmark Pathways upregulated in each subgroup





Differential super-enhancer landscapes across samples identifies normal- and tumor-specific transcriptional circuitry



Name Transcription Factor ○ Non-Transcription Factor Gene

Ovarian Cancer subtype specific

Shared between normal and ovarian cancer sub-

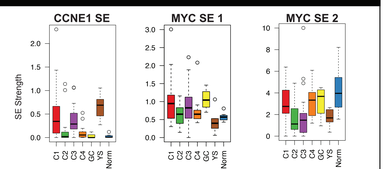
Normal tissue specific

Significant SE-SE mutual information

A mutual information network of subgroup- and normal-specific SEs was constructed to demonstrate the complex circuitry underlying ovarian cancer. Rectangles represent SEs linked to TF transcripts. Circles represent SEs linked to non-TF transcripts. Edges represent mutual information relationships stronger than a stringent cutoff based on a shuffled empirical background. All SEs depicted here were identified as specific to either a patient subgroup or normal tissue. The SEs present in both patient tumor subgroups and normal are depicted in gray, while those SEs present only in tumor samples are shown in red.

Some cluster-specific SEs suggest increased tumor dependency on transcriptional machinery

Cluster-specific SEs imply that the tumors falling into those clusters have an increased transcriptional dependency on active transcription initiation at these loci, as well as the components of the corresponding transcriptional machinery. Three examples shown here (right) demonstrate cluster-specific SEs at CCNE1, and two cluster-specific SEs at MYC.



Conclusions

We have assembled the largest H3K27Ac ChIP-seq database in ovarian cancer and related normal tissue to date, with the express purpose of profiling active enhancer activity. We complemented our ChIP-seq data with RNA-seq and mutational profiling. We were able to show large differences between the enhancer landscapes of tumor and normal in ovarian cancer. We were also able to identify 6 total patient subgroups, comprising granulosa cell, yolk sac, and 4 novel subgroups with samples from clear cell, high grade serous, low grade serous, mucinous, and endometrioid ovarian cancer. Each of these subgroups is driven by its own unique transcriptional circuitry, with associated transcription factors and potential druggable targets. Interestingly, while we can predict subgroup membership for cell lines, we find that many cell lines have a transcriptional signature distinct from patients (volcano plot below, showing thousands of genes significantly differentially expressed in cell lines). Many genes involved in ovarian cancer etiology are linked to cluster-specific SEs, such as CCNE1, MYC, and CD47 (above and left). Many other potential targets are revealed by our analysis.