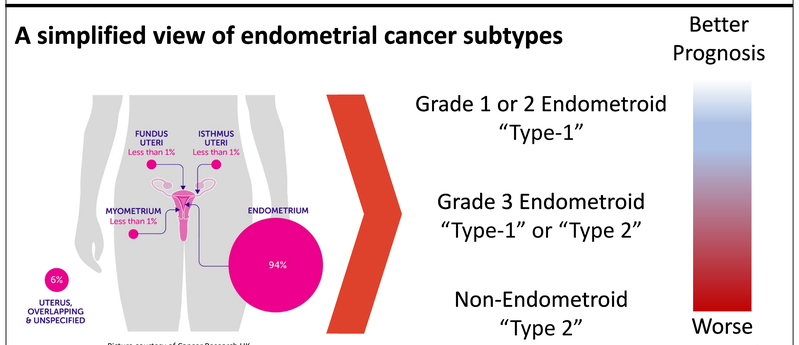
Clustering from structural variation in endometrial cancer

Julian Stanley, Shreepriya Das, Michele Cummings, Nicolas Orsi, & Jeremy Gunawardena

Department of Systems Biology, Harvard Medical School, Boston, MA 02115

Endometrial cancers are the fourth most common cancer in women worldwide, and identification of molecular characteristics that define subtypes of the cancer may complement traditional diagnostic techniques. We obtained 95 hysterectomies of different grades of endometrial cancer and analyzed structural variation events from formalin-fixed samples. Hierarchical clustering of copy number variation (CNV) revealed two clusters of endometrial cancer that also correspond to survival. Clustering can also be accomplished by a simple technique of plotting patients in terms of aggregate copy number gain and loss. Further work needs to be done to identify genetic markers that may aid diagnosis, and how those markers compare across cancer types.

A simplified view of endometrial cancer subtypes



Better Prognosis

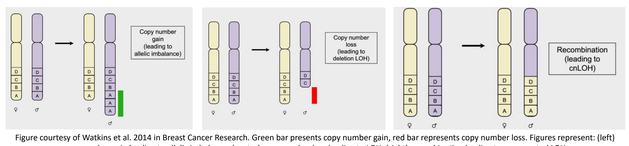
Grade 1 or 2 Endometroid "Type-1"

Grade 3 Endometroid "Type-1" or "Type 2"

Non-Endometroid "Type 2"

Worse Prognosis

Copy number variation (CNV), allelic imbalance, and loss of heterozygosity (LOH) are types of structural variation



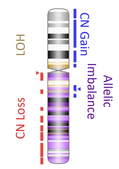
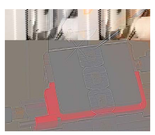


Figure courtesy of Watkins et al. 2014 in Breast Cancer Research. Green bar presents copy number gain, red bar represents copy number loss. Figures represent: (left) copy number gain leading to allelic imbalance, (center) copy number loss leading to LOH, (right) recombination leading to copy neutral LOH.

Structural variation analysis



Picture courtesy of Leica Biosystems

Picture courtesy of Affymetrix

DNA extracted from FFPE tissue was analyzed using the Oncoscan Copy Number assay

Better Prognosis

Grade 1 Endometroid (11)

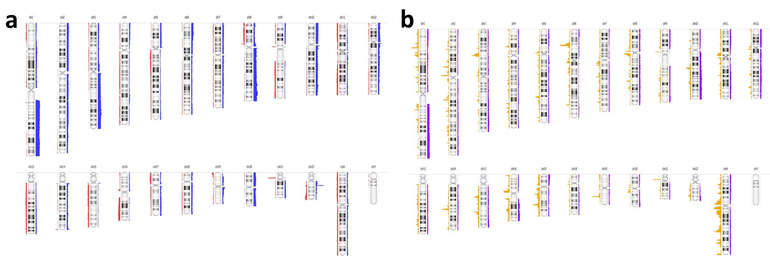
Grade 2 Endometroid (21)

Grade 3 Endometroid (32)

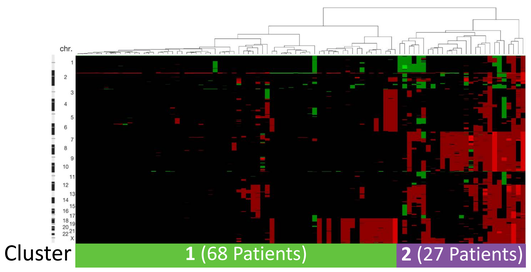
Non-Endometroid (31)

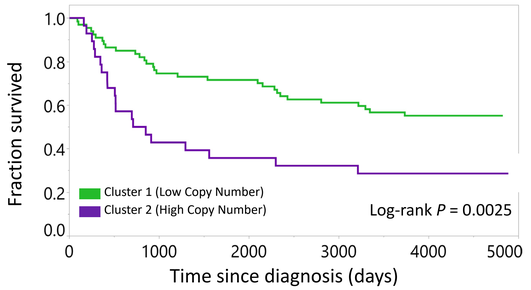
Worse Prognosis

Figure 1. Aggregate **(a)** copy number gain and loss and **(b)** allelic imbalance and loss of heterozygosity in 95 endometrial cancer FFPE samples



CNV-based clusters





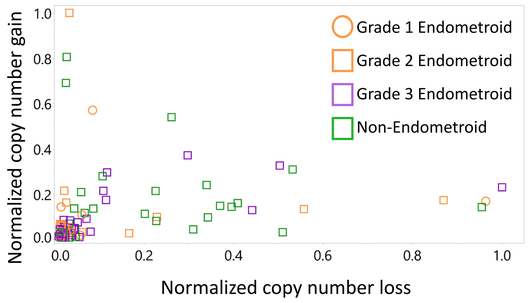


Figure 2.

Method: Unsupervised hierarchical clustering based on Euclidean distance, Ward linkage, and silhouette optimization.

a: Copy number variation separates patients into two clusters. Green represents CN loss and red represents CN gain.

b: Clusters have a statistically and clinically significant difference in survival.

Figure 3. Patient clusters in copy number space, labeled by clinical diagnosis.

K-means clustering on a simple metric shows clusters that have significant difference in survival (not shown).

Does the method also apply to ovarian cancer?

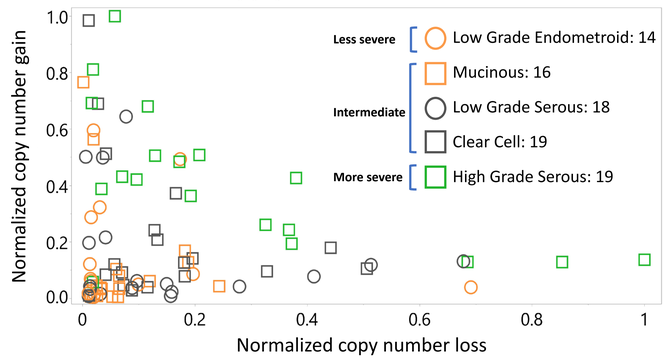


Figure 4. Ovarian cancer patient clusters in copy number space.

- ❖ Low grade endometroid are least severe → Low copy number
- ❖ Mucinous is intermediate, but based on CN we predict less severe
- ❖ Low grade serous and clear cell are intermediate; they distribute normally
- ❖ High grade serous are most serious → High copy number

Moving forward: comparative analyses

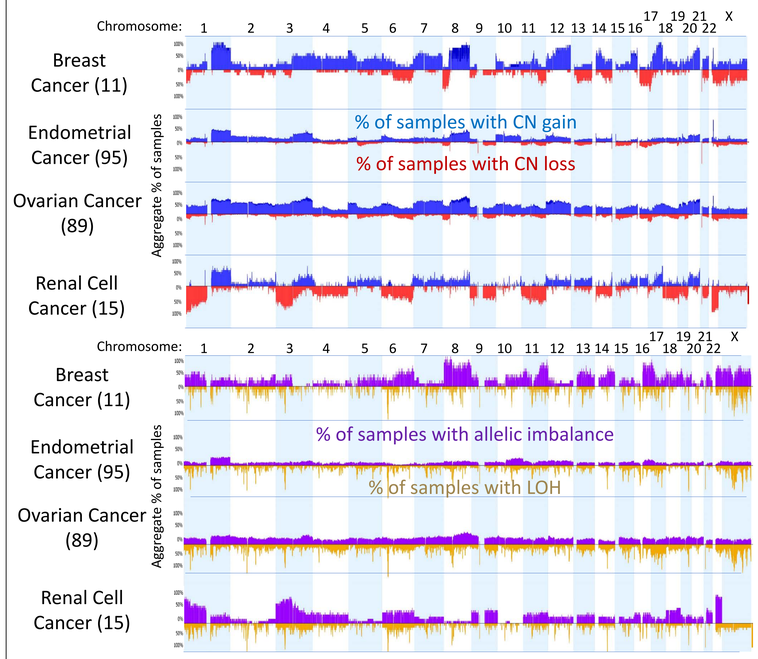


Figure 5. Copy number calls were made and analyzed by Nexus Express Software. Breast and renal cell cancers were used as comparisons with respectively similar and different risk factors as endometrial and ovarian cancer. **a:** CN gain (blue) and CN loss (red) and **b:** allelic imbalance (purple) and LOH (brown) shown as a proportion of samples with the given variation.

Discussion

Both endometrial and ovarian cancer are extremely heterogenous diseases that affect hundreds of thousands of women worldwide. Microarray analysis is a powerful tool that may provide information that can lead to more personalized cancer treatments.

The Oncoscan platform, based on the molecular inversion probe technique, can consistently detect variation heavily-degraded FFPE clinical samples.

Clustering techniques and comparative analyses are necessary to understand what differences between cancers are visible from genomic analyses, and how different molecular profiles can complement traditional diagnostics.

Acknowledgements

I am extremely grateful to HMS Systems Biology and the Summer Internship in Systems Biology program for the welcoming working environment. Thanks to Shreepriya Das and Jeremy Gunawardena for their invaluable mentorship during this short internship, and to Nicolas Orsi and Michele Cummings for providing samples and helpful clinical expertise.

References

1. Levine, DA et al. 2013 Integrated genomic characterization of endometrial carcinoma. Nature 497, 67-73.
2. Zhang, J. 2018 CNTools: Convert segment data into a region by sample matrix to allow for other high level computational analyses. R package version 1.36.0.
3. Kaveh, F. et al. 2016 A systematic comparison of copy number alterations in four types of female cancer. BMC Cancer 16:913.
4. Jung, H. et al. 2017 Utilization of the Oncoscan microarray assay in cancer diagnostics. Applied Cancer Research 37:1.

Contact: julianstanley@fas.harvard.edu