



JOHNS HOPKINS
M E D I C I N E

DeepTCR: A Deep Learning Framework For Revealing Structural Concepts within TCR Repertoire

John-William Sidhom, MD.,PhD. Candidate

Drew M. Pardoll, MD.,PhD.

Alexander S. Baras, MD.,PhD.

No Disclosures

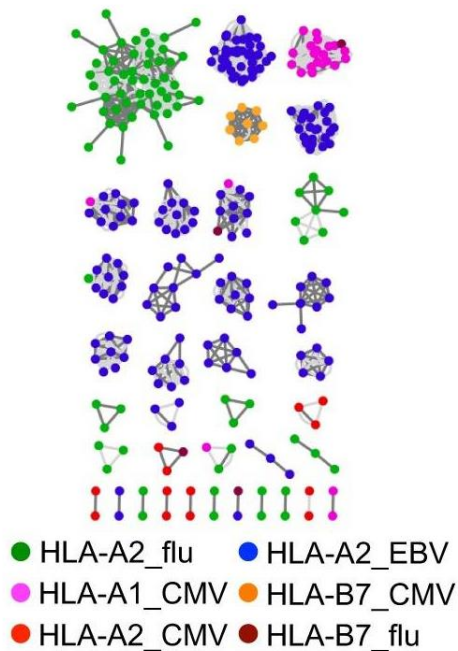
T-Cell Receptor Sequencing

- Count-based NGS method of CDR3
- Quantitative assessment of T-Cell repertoire
- Sequence acts as 'barcode'

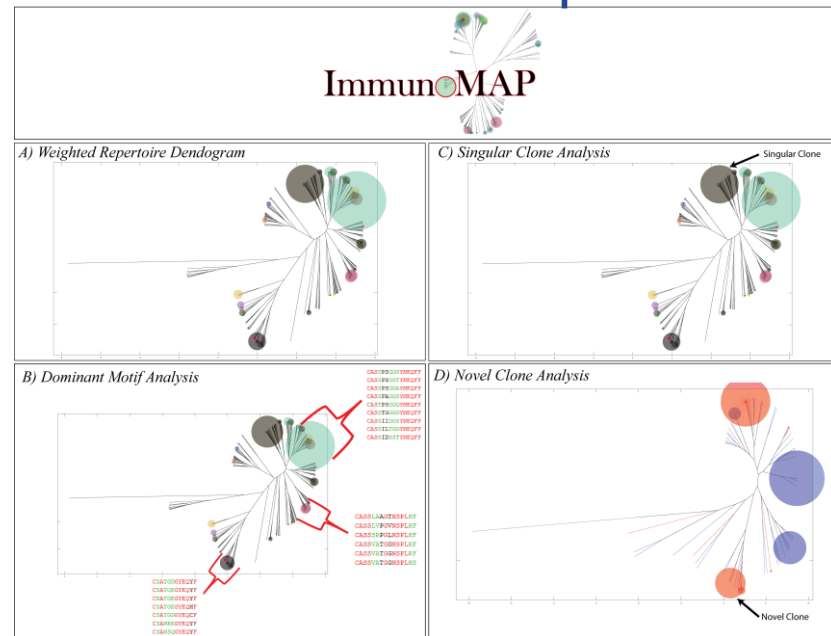
nucleotide	aminoAcid	count (templates/reads)	frequencyCount (%)
TTCTCCCTCATTCTGGAG	CASGTGDNQAPLF	25	0.030194623
GAGAACTTCTCCCTCATT	CASGDPDTQYF	22	0.027358645
CTCACTGTGACATCTGCC	CASSIKQGQFDEQYF	15	0.021143412
TTCTCTCTCATTCTGGAG	CASSAQGAGEQYF	12	0.01495893
GAGAACTTCTCCCTCATT	CASGTGDTQYF	11	0.013469187
CTGCTGGAATTGGCTTC	CASSVTGGANTGQLYF	10	0.01300108
TATTTCACTCTGAAAATC	CASSLGQYEQYF	10	0.013209507
TTCTCTCTGCTGGAATTC	CASSDSYNNQAPLF	10	0.013209507
AATCTTCGAATCAAGTCT	CASSPPGLGETLYF	10	0.014521574
TTCTCCCTCATTCTGGAG	CASGAGDYAEQFF	9	0.012857572
CTCCTGCTGGAATTGGC	CASSDGGASTGQLYF	9	0.010387196
CTCCTGCTGGAATTGGC	CASSDVGGPYAEQFF	9	0.00955007
GAGATGAACATGAGTGC	CASSPTGGAPEQYF	8	0.011439583
CACTCTGAAGATTCAACCTACAGAACCCAAGGACT		8	0.008689026
CAGCCTAGAAATTTCAGT	LCQQSVTGGAEQYF	8	0.012119535
TTCTCTCTGCTGGAATTC	CASSDSYNNQAPLF	8	0.010732297
AATCTCTCTGCTGGAATTC	CASSDSYNNQAPLF	8	0.010732297

Current Structural Approaches

GLIPH¹



ImmunoMap²



¹Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L. E., Rubelt, F., ... & Haas, N. (2017). Identifying specificity groups in the T cell receptor repertoire. *Nature*, 547(7661), 94.

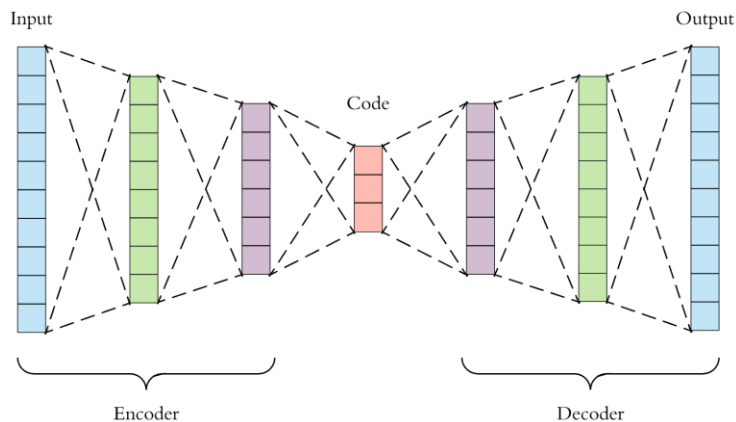
²Sidhom, J. W., Bessell, C. A., Havel, J. J., Kosmides, A., Chan, T. A., & Schneck, J. P. (2018). ImmunoMap: A Bioinformatics Tool for T-cell Repertoire Analysis. *Cancer immunology research*, 6(2), 151-162

DeepTCR

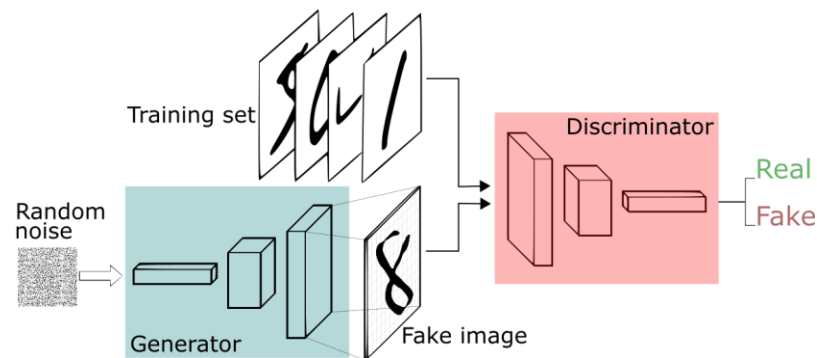
- Unsupervised and Supervised Deep Learning Methods to Parse TCRSeq
- Unsupervised Methods:
 - Uncovering structure in data when labels are not provided
 - Clustering Homologous Sequences
 - Structurally Comparing Whole Repertoires/Samples
- Supervised Methods:
 - Using labeled data to guide training process.
 - Discovering predictive structural signatures in the data

Unsupervised Methods

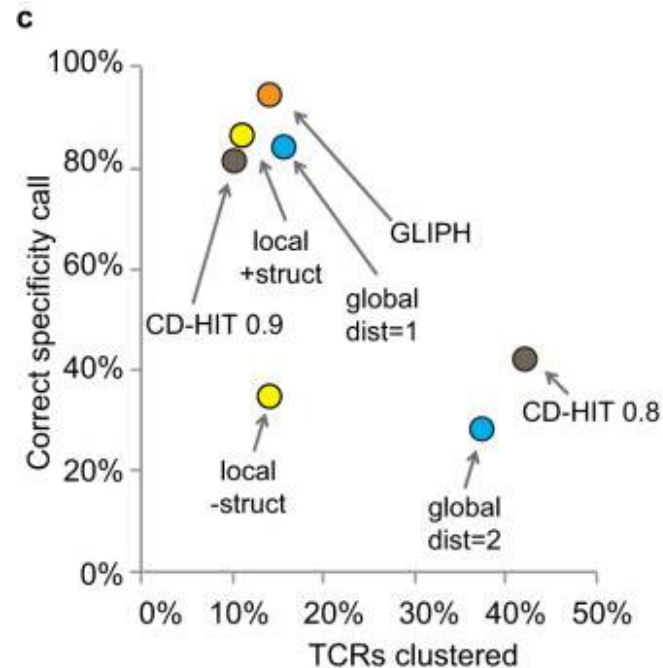
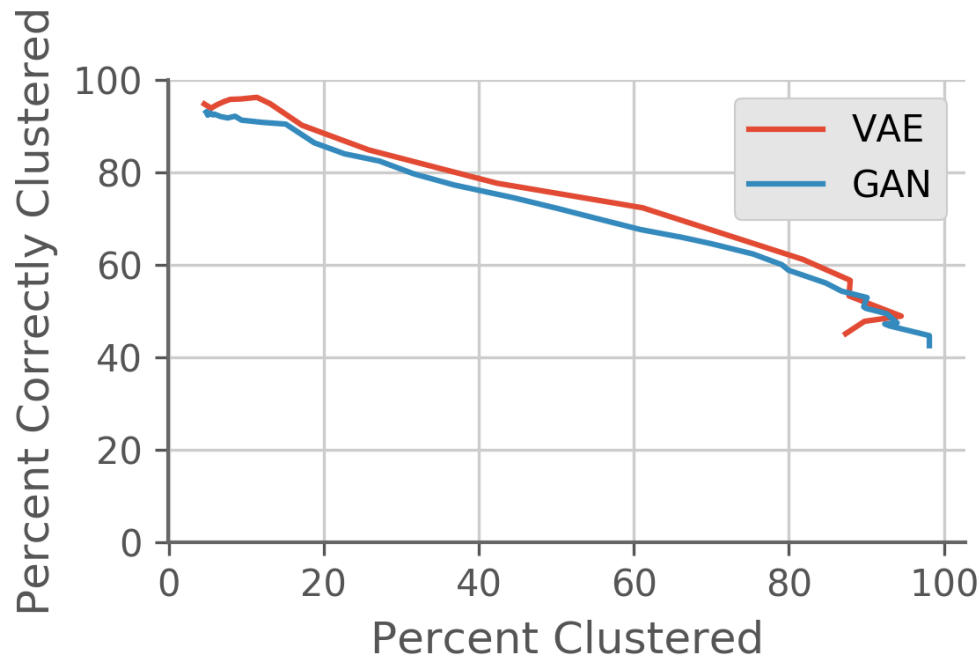
Variational Autoencoder (VAE)



Generative Adversarial Network (GAN)

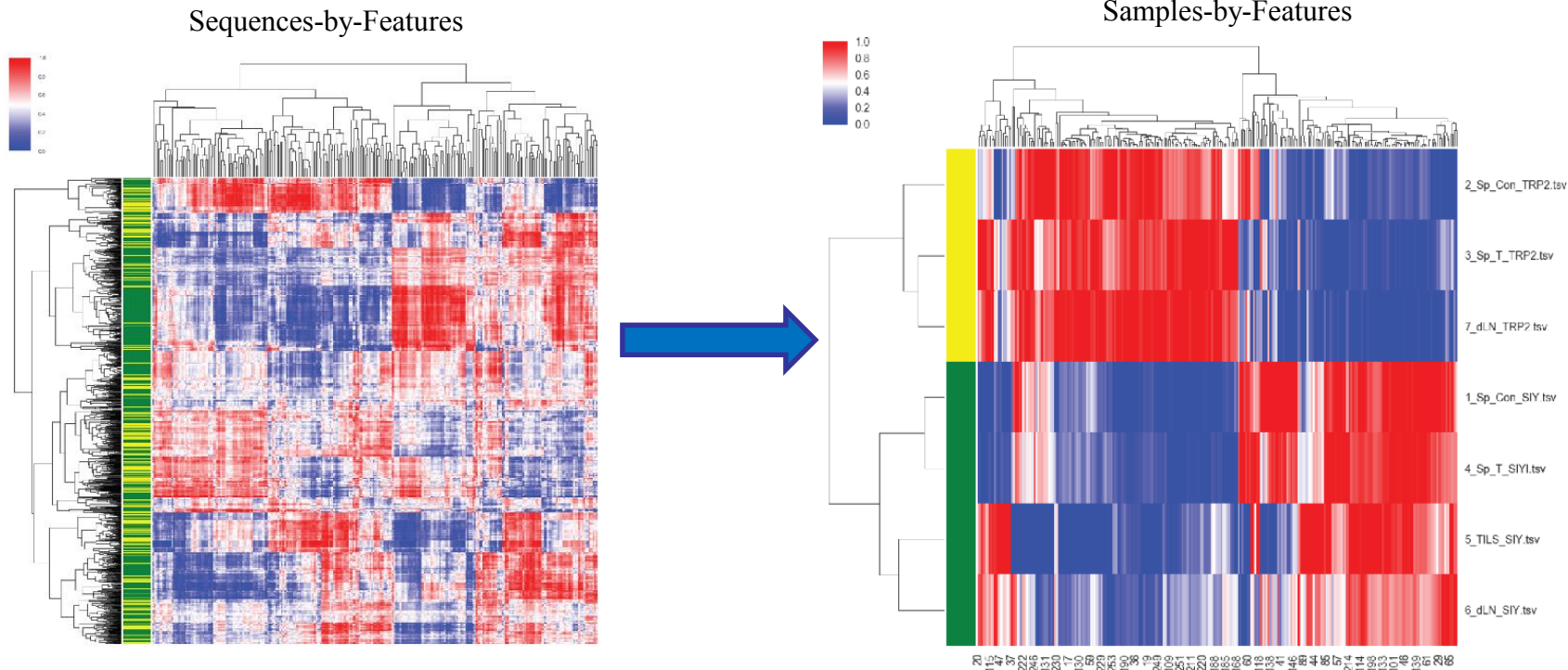


Benchmarking against GLIPH



Wholistic Structural Analysis

SIY/TRP2



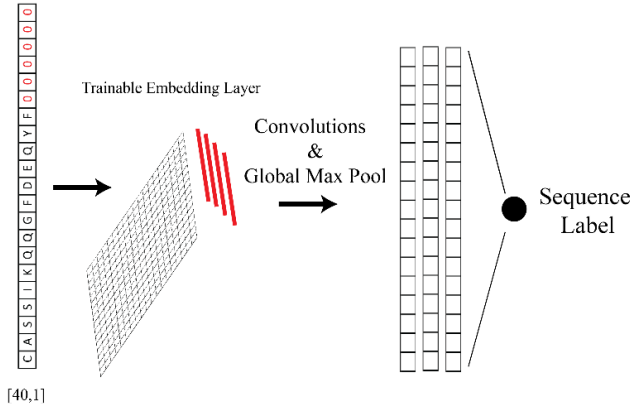
Key Advantages

- Uses TensorFlow that can take advantage of GPU's, thus able to process millions of TCR sequences within seconds.
- Requires ***only*** β -chain sequence.
- Structural comparisons of whole repertoires, not just sequences

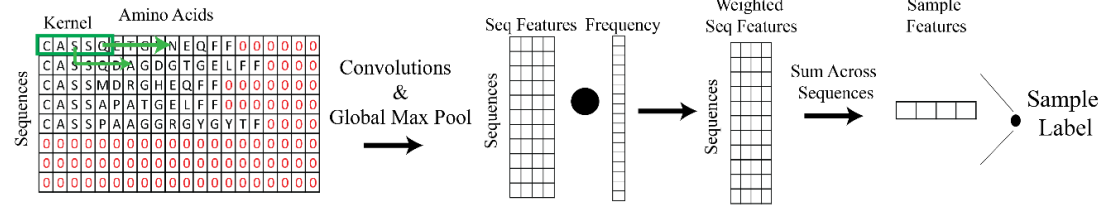
Supervised Methods

Supervised Deep Learning Architectures

C *Single Sequence Classifier*

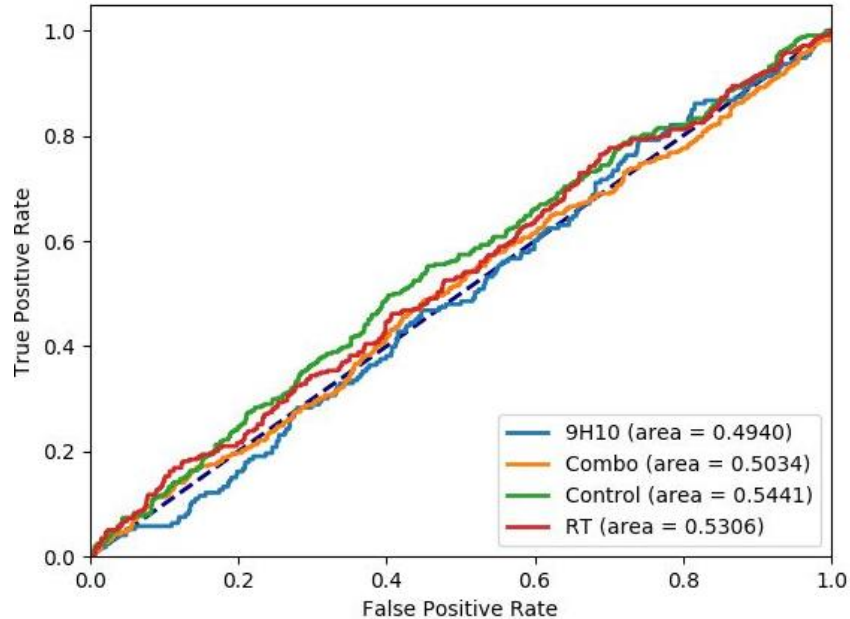


d *Whole Sample Classifier*

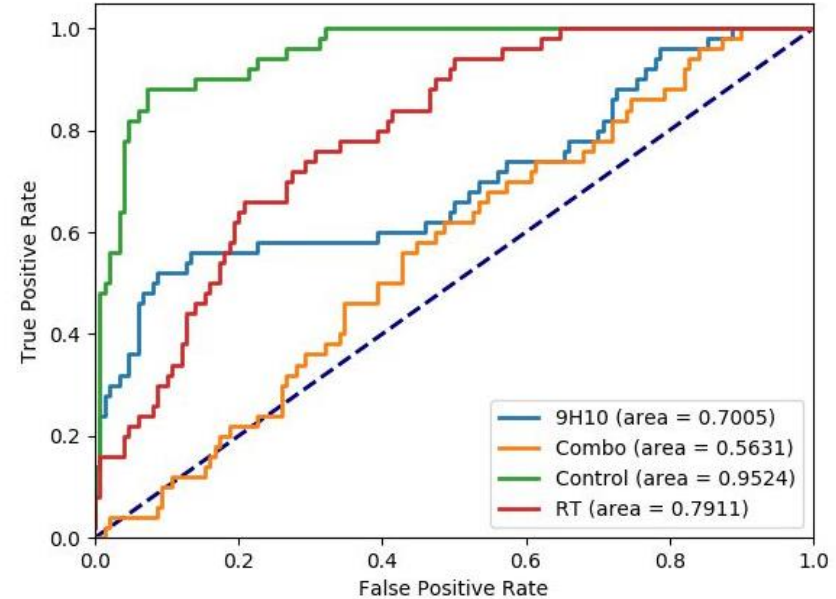


TIL Repertoire - Immunotherapy

Single-Sequence Classifier



Whole Sample Classifier

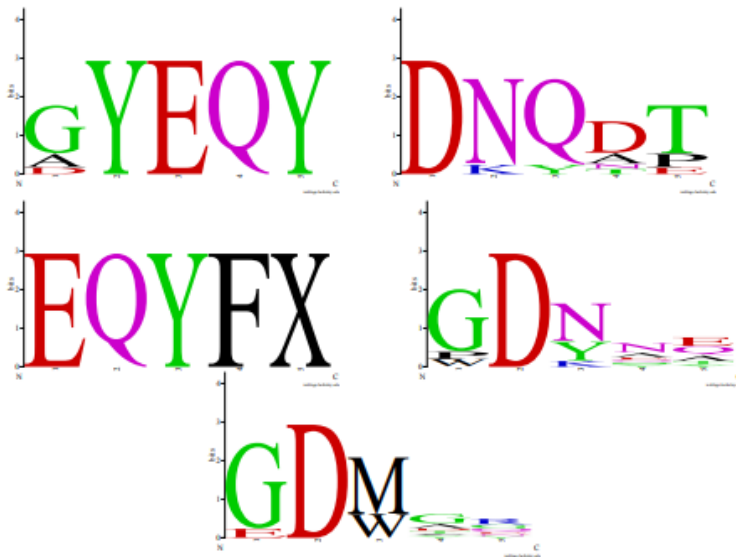


Rudqvist, N. P., Pilonis, K. A., Lhuillier, C., Wennerberg, E., Sidhom, J. W., Emerson, R. O., ... & Demaria, S. (2018). Radiotherapy and CTLA-4 blockade shape the TCR repertoire of tumor-infiltrating T cells. *Cancer Immunol Res*, 6, 139-50.

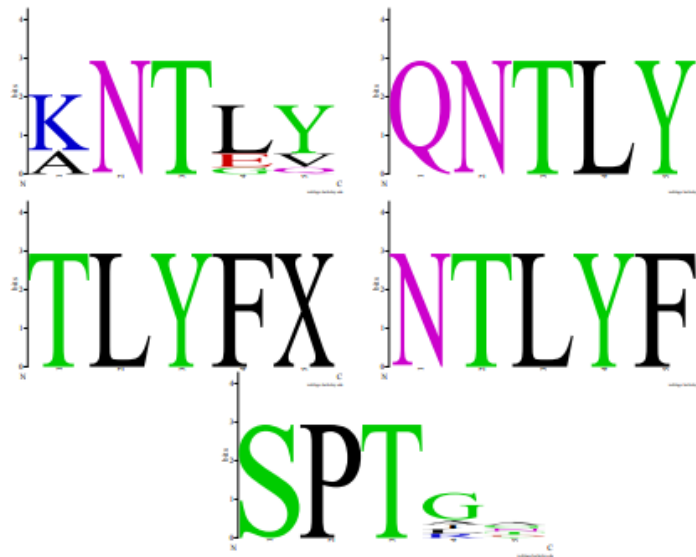
‘Explainability’ of Neural Networks

e

Control Specific Motifs



RT Specific Motifs



How can you use DeepTCR?

- Pre-Print @ Biorxiv
- Python Package & Github Repository
 - <https://github.com/sidhomj/DeepTCR>
 - Source Code/Easy Install Instructions
 - Jupyter Notebook Tutorial
- Questions?
 - jsidhom1@jhmi.edu
 - Poster: P588

```
from DeepTCR.DeepTCR_U import DeepTCR_U

"""Unsupervised Analysis"""
DTCRU = DeepTCR_U('Rudqvist_U')
DTCRU.Get_Data(directory='Rudqvist/', Load_Prev_Data=False, type_of_data_cut='Num_Seq', data_cut=20)
color_dict = {'Control': 'y', 'RT': 'b', '9H10': 'g', 'Combo': 'm'}

#Train VAE and output heatmaps/clusters
DTCRU.Train_VAE(accuracy_min=0.8, Load_Prev_Data=True, latent_dim=256)
DTCRU.HeatMap_Sequences(filename='Rudqvist_VAE_HM_Sequences.tif', color_dict=color_dict)
DTCRU.HeatMap_Files(filename='Rudqvist_VAE_HM_Files.tif', color_dict=color_dict)

# #Train GAN and output heatmaps/clusters
DTCRU.Train_GAN(Load_Prev_Data=False)
DTCRU.HeatMap_Sequences(filename='Rudqvist_GAN_HM_Sequences.tif', color_dict=color_dict)
DTCRU.HeatMap_Files(filename='Rudqvist_GAN_HM_Files.tif', color_dict=color_dict)

#Cluster Sequences
DTCRU.Cluster(write_to_sheets=True)
```