UNIVERSITY OF CALIFORNIA SAN DIEGO

Elucidating the mechanism by which aneuploidy interferes with immune response during tumor progression

A thesis submitted in partial satisfaction of the requirements of the degree

Master of Science

in

Bioengineering

by

Su Xian

Committee in charge:

Professor Hannah Carter, Chair
Professor Yingxiao Wang, Co-Chair
Professor Stephanie Fraley

2018

The Thesis of Su Xian is approved and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank UCSD faculty and staff for caring about their students and putting effort into helping them. And to Brian Tsui in Carter lab who generously shared expression data profiles with all lab members, as well as for sharing of his knowledge of all kinds of genome library information. Thanks to Michelle Dow in Carter lab for her patient instructions guiding me through the basic knowledge of computing, and her encouragement. Thanks for Rachel Marty for her always generously helping my writing and presentation skills and always being patience when my project came to problems. Thanks for Kivilcim Ozturk in kindly sharing presentation skills and improved my spoken ability. And thanks to all the members in Carter lab for supporting each other. Thanks for our collaborator Maurizio Zanetti and Jeffery Rodvold for providing biological details and careful guidance in this project. Without all those helps I would not be able to have this project in progress so fluently and would not be able to stand and speak out my idea in a compelling and confident way. Finally, I would like to express my gratitude to Professor Hannah Carter, for letting me, an unexperienced student who had absolutely zero computational background, conduct independent research under her guidance. Without her ceaseless encouragment and inspiration I would be lost in pursuing my research interests and future career.

ABSTRACT OF THE THESIS

Elucidating the mechanism by which aneuploidy interferes with immune response during tumor progression

by

Su Xian

Master of Science in Bioengineering

University of California, San Diego, 2018

Professor Hannah Carter, Chair
Professor Yingxiao Wang, Co-Chair

Cancer cells frequently carry chromosomal abnormalities, called aneuploidy. Aneuploidy has been reported to have association with impaired immune response in tumors. Various studies focus on investigating aneuploidy in cancer to gain insights into potential therapeutics for cancer. Endoplasmic reticulum (ER) stress, which is as prevalent as aneuploidy in tumor cells, has also been suggested to have potential impact on immune surveillance during tumor progression. We have experimentally validated that drug induced aneuploidy in vitro leads to an increase of ER stress in cells, which will then initiate the unfolded protein response (UPR). In this thesis we provide a new perspective, investigating ER stress as the potential mechanism by which aneuploidy helps tumor cells escape from immune response during progression. We

confirmed the activation signal from 3 branches of UPR and among which, PERK remained universally unperturbed across 31 cancer types. The activity of PERK downstream transcription factor, EIF2A and DDIT3, were highly correlated with aneuploidy level, suggesting cross talk between the PERK branch pathway and aneuploidy state. Partial correlation testing and linear models were applied to confirm the immune-suppressive function of EIF2α and DDIT3. These results support that the PERK pathway could potentially be a new therapeutic target in cancer treatment.

Chapter 1 Introduction

1.1 Aneuploidy

Aneuploidy, a term that used to describe chromosome abnormality, depicts a phenomenon where the cell does not maintain the correct number of chromosomes compared with healthy cells [1]. For example, a normal human cell should have 46 chromosomes, while on average colorectal cancer cells are reported to have 59 chromosomes [2].

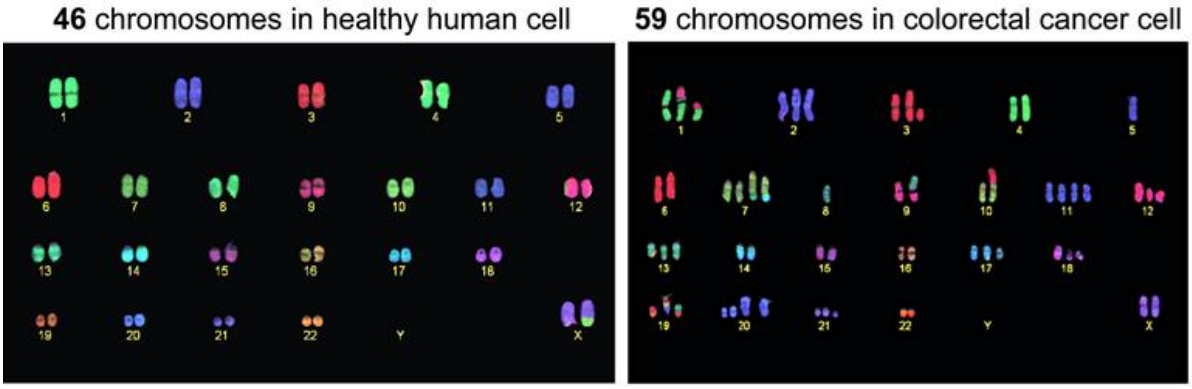


Figure 1. Example of aneuploidy. A healthy human cell karyotype with 46 chromosome numbers (left). In comparison, a colorectal cancer cell karyotype with 59 chromosomes (right), illustrating a frequent characteristic of cancer cells, namely aneuploidy.

The presence of aneuploidy could be caused during cell divisions due to the inappropriate separation of chromosomes resulting in unbalanced chromosome numbers in daughter cells [1]. Both loss and gain of chromosome copy numbers are considered aneuploidy. Interestingly, aneuploidy is fairly prevalent among various cancers. Previous studies also showed that aneuploidy plays role in tumorigenesis [3]. *Davoli et al* found that aneuploidy could impair immune responses in tumor cells to

promote their survive. According to *Davoli et al*, high levels of aneuploidy are correlated with poor immune responses. However, the mechanism by which aneuploidy could interact with immune reactions has not been revealed.

During cancer cell progression, the presentation of aneuploidy is also accompanied by cellular stress, including proteotoxic stress, oxidative stress, etc. [4]. Proteotoxic stress will produce a large amount of protein that is miss-folded due to the increase of folding demand that overwhelms the endoplasmic reticulum (ER) loading capacity. In this case ER stress is initiated.

Zanetti et al describe evidence that in a scenario of aneuploidy induced endoplasmic reticulum overload, ER stress can result in the generation of a cell-nonautonomous factor that acts to suppress anti-tumor immunity [5]. The ER represents an early step in the process by which cells export functional proteins to sustain cellular homeostasis. Once the balance of protein production and modification has been broken, cells will suffer various functional abnormalities including immune deficiency due to a lack of production of major histocompatibility complex class I (MHCI) molecules in T cells [6]. Under ER stress, the cells themselves are trying to return to homeostasis, thus a rescue pathway called the unfolded protein response (UPR) is activated to restore the proper function of the ER [7].

The UPR has three downstream branch pathways. Usually, in the presence of unfolded protein or miss-folded protein, the ER will activate the UPR to produce more chaperones to assist protein-folding, and simultaneously supress transcription levels to reduce the ER load [7]. The Immunoglobulin Heavy Chain-Binding Protein (Bip), is a key regulator of the UPR. During normal conditions, the UPR is inactivate and Bip binds

with 3 sensor proteins called IRE1α, ATF6 and PERK. However, due to the higher binding affinity between Bip and unfolded or miss-folded proteins, in the presence of unfolded and miss-folded proteins, Bip will bind with unfolded or miss-folded proteins to assist the folding. When Bip leaves the 3 sensors, IRE1α, PERK and ATF6 will then induce the downstream pathway to start the “rescue” of ER function [7].

IRE1α auto-phosphorylates itself, oligomerizes and activates its endoribonuclease function. This will generate a splice isoform of X-box binding protein-1 (XBP1s), which acts as a transcriptional regulator to produce various ER chaperones that promote proper ER protein folding activity [7].

PERK homodimerization and autophosphorylation activates its cytoplasmic kinase domain which can phosphorylate eukaryotic initiation factor 2α (EIF2α). EIF2α acts as a regulator that reduces global mRNA translation. This helps to attenuate the amount of protein entering the ER to further resolve the stress [7].

ATF6 translocate to the Golgi apparatus and is cleaved by Golgi enzyme site 1 proteases (S1P and S2P), which converts it into its functional form. It then travels to the nucleus and induces genes that can increase the capacity of the ER to help alleviate ER stress [7].

Zanetti *et al.* found that when co-cultured with high ER stress cancer cells, dendritic cells present in an inactivated state that cannot generate a robust immune response [5]. Thus, ER stress could potentially play roles in tumor growth and impaired immune responses. In several previous studies and reviews, induction of the UPR was also associated with a reduction in major histocompatibility complex I (MHCI) production, and a significantly increased pro-inflammatory response [5]. Thus both

aneuploidy and ER stress are show evidence of a correlation with impaired immune response. This lead us hypothesize that genomic instability and ER stress interact with each other to help cancer cells evade immune surveillance.

1.2 The Bridge Between Aneuploidy And Immune Response

As previously mentioned, an impaired immune phenotype was witnessed in high aneuploidy tumors in a previous study [3]. While it is not clear how aneuploidy, the change in copy number of chromosomes, causes an escape from immune response in cancer cells. Here we posed this question from a new perspective, via the careful study of ER stress. Aneuploidy is evidenced to have association with various cell stresses including proteotoxic stress which would cause an overload of the ER lumen and induce the ER stress [5]. ER stress will activate a rescue response called the UPR, which is been reported to have association with impaired immune response [5]. For example, the artificially induced accumulation of proteins in B cells resulted in a UPR and upregulation of MHC class II levels. However, the cells displayed fewer high-affinity peptide/MHC class II complexes [8]. Moreover, in mouse thymoma, palmitate or glucose deprivation induced ER stress was found to down-regulate MHC class I antigen presentation [6].

In this scenario, we posit that, ER stress induced UPR could be the bridge that connects the impaired immune response with aneuploidy, motivating investigation of different aspects of ER stress to identify mechanisms by which aneuploidy interferes with immune response.

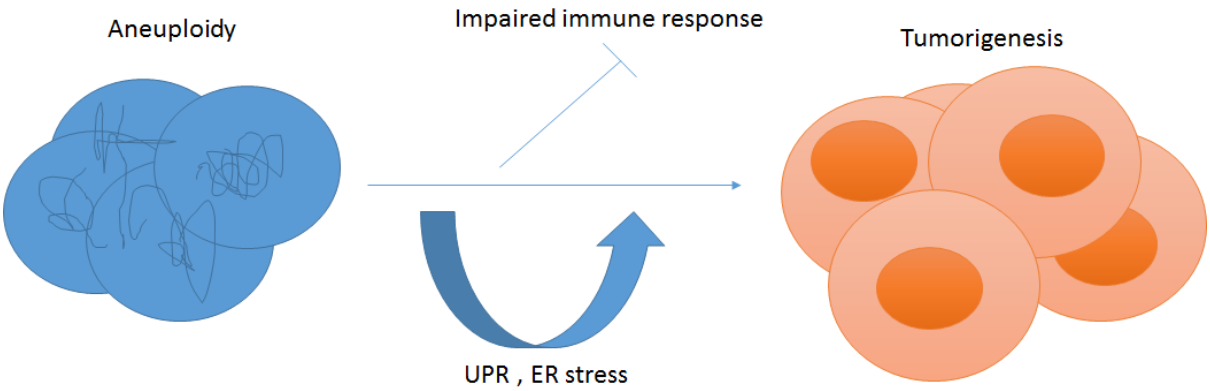


Figure 2. The main hypothesis. Aneuploidy, through its interaction with ER stress and the UPR pathway, disturbs the aniti-tumor immune response which results in tumor progression.

Chapter 2 Methods

2.1 Quantifying Aneuploidy

In the quantification of aneuploidy, we downloaded the processed Affymetrix SNP6 data from The Cancer Genome Atlas (TCGA). Each processed Affymetrix SNP6 data would have a normal sample and tumor sample from the same patient. The processed data provides information of copy numbers quantity for every small fractions of chromosome, which is called segmentations. In the processing methods, the normal sample was used as a reference, and the relative amplitude of segmentation fold change is calculated between tumor and normal samples to represent the copy number change level. The rest of the analysis was based on the relative fold change between tumor and normal samples in order to quantify aneuploidy.

In this study, we defined the somatic copy number alteration (SCNA) scores for each patient. These scores cover three types of copy number aberration: focal level, arm level and chromosome level events. We applied a signal cutoff of 0.1 fold change in both directions (meaning gain or loss) according to *Beroukhim et al*. [8].

For each chromosome, the focal level describes copy number changes that affect a very small fraction of a chromosome (less than an arm level, either chromosome p arm or chromosome q arm, are quantified as focal level SCNA), which is also considered as the result of many selection events during evolution of the cancer genome [9]. Thus, a segment in the processed Affymetrix SNP6 data that spans a length less than either the chromosome p arm or q arm satisfying the 0.1 signal cutoff would be quantified as a focal SCNA event. For instance, a small fraction of

chromosome 1q amplified with a fold change of 0.3 would be quantified as 1 focal level gain.

Arm level SCNAs are copy number changes that span an entire chromosomal arm (either chromosome q arm or p arm). Thus, in the processed Affymetrix data, all the segmentations within a chromosome p arm or q arm that sharing the same direction of fold change that satisfying the threshold would be quantified as an Arm SCNA event. For example, all segments within chromosome 1q shows a fold change less than -0.1 would be quantified as 1 arm level loss.

Whole chromosome deletion or duplication is counted as a chromosome level SCNA which means for all segments within this chromosome, they all satisfy the threshold in the same direction.

In total, the collected Affymetrix SNP6 array data from TCGA have 8268 samples including 31 cancer types. After quantifying the 3 types of SCNA events for each patient, we summed up the number of events that observed (including both gain and loss) within each type of SCNA as the corresponding SCNA score. For example, for each patient j with i numbers of observation of events happening in focal, arm and chromosome SCNA, the focal SCNA, arm SCNA and chromosome SCNA are calculated using the formula:

$$focal\ SCNAj = \sum_{i} |\ focal\ copy\ number_i\ |$$

$$Arm\ SCNAj = \sum_{i} |\ Arm\ copy\ number_i\ |$$

$$Chromosome\ SCNA_j = \sum_i |\ Chromosome\ copy\ number_i\ |$$

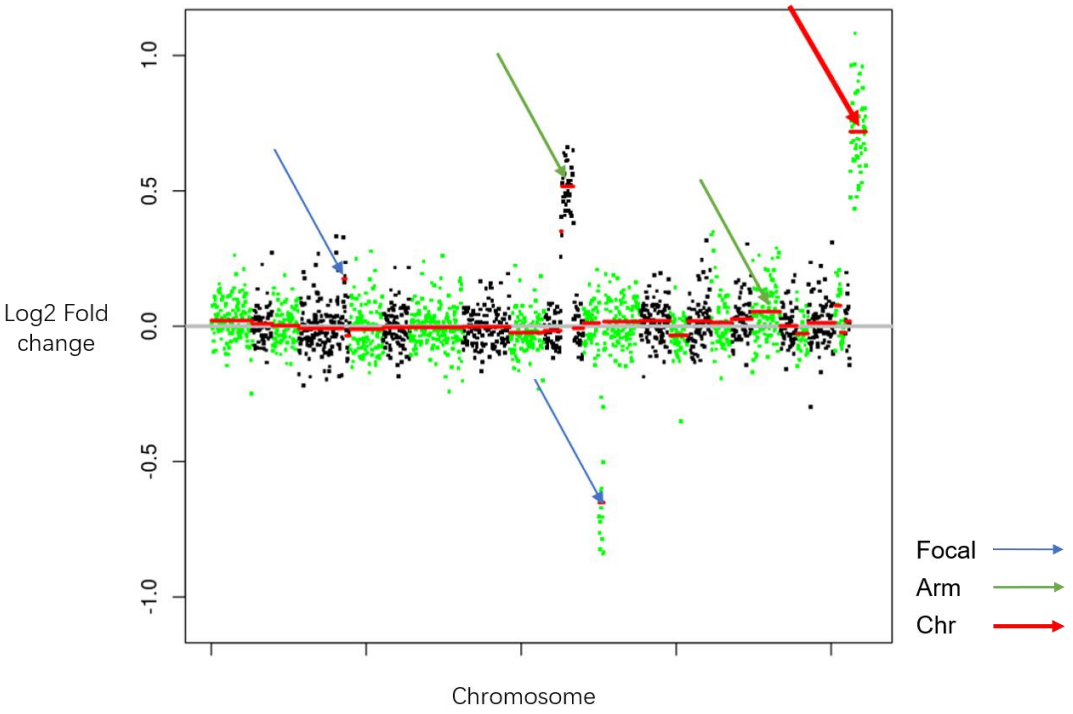


Figure 3. Example of quantification of SCNAs. In this figure, each green and black section corresponds to a different chromosome. Chromosomes are sorted from 1 to 22 from left two right, which consecutive chromosomes alternating in color. The y-axis represents the $\log_2$ fold change of SNP marker copy number. The red line shows the mean value of copy number for each chromosome arm or segment. Blue arrows indicate focal SCNAs; green arrows indicate arm level SCNAs; red arrow indicates a chromosome level SCNA. For the amplification and deletion threshold, we set a threshold of 0.1 fold change in either direction (0.1 or -0.1).

Then for the subsequent study, we generated a single score by adding up z-scored focal, arm and chromosome level SCNA counts for every sample according to the following formula:

$$A_i = zscore(focal_i) + zscore(arm_i) + zscore(chr_i),$$

where the $zscore$ is calculated by:

$$zscore(x) = \frac{(x - x_{mean})}{std(x)}$$

The $std(x)$ is the standard deviation of x in the distribution. In this way A_i generally represents the total aneuploidy level resulting from 3 features (focal, arm and chromosome). This formula is applied the same from *Davoli et al* [3].

Apart from the A_i score, we also generated a total SCNA score for every individual sample by the formula:

$$SCNA_i = scaled(focal_i) + scaled(arm_i) + scaled(chr_i).$$

The $scaled$ is calculated by

$$scaled(x) = 10 \cdot \frac{(x_i - x_{mean})}{(x_{max} - x_{min})}$$

Under this formula, 3 features are scaled in the same range and sum up to be the SCNA score. The SCNA score is more useful when applying correlation study or regression models because it does not have a positive negative cancelling out problems like the A_i score we calculated above.

2.2 Gene Expression Data and statistical test

In order to test the ER stress gene expression, we collected the RNA-seq from TCGA and reprocessed it. The gene expression data was reprocessed by Brian Tsui in Carter lab, who applied a tool called sailfish, generated an expression matrix containing 10989 samples, 27105 genes, in total 32 cancer types. Genes that playing important roles in the ER stress response according to previous study are selected out, including ERN1, ATF6, EIF2AK3, EIF2α, ATF4, XBP1, HSPA5, HSP90B1, MAP3K5, MAPK8, TRAF2, DDIT3, DNAJB9, CYCS, MFN2, PPP1R15A. Gene expression data was quantified in Transcripts Per Million (TPM). Log 2 value was taken to reshape gene expression to normal distribution for the subsequent statistic test. Samples with mutations in selected genes are been removed to ensure the gene network integrity. For every samples in tumor types, only primary tumor site are taken into analysis. Normal samples in each corresponding cancer types were used in statistical analysis in comparison with tumor samples. After filtering samples mentioned above, the total size of analysis were then conducted based on 8473 tumor samples and 712 normal samples.

In order to test if ER stress is presenting among tumor, we applied statistic t-test to the expression level differences of selected ER stress genes between our tumor samples and matched normal samples. The p value is adjusted using benjamini-hochberg methods.

2.3 Gene Set Enrichment Analysis (GSEA)

After collecting of Expression data and confirming the up-regulation of HPSA5 among several cancer types, we applied a tool called Gene Set Enrichment Analysis

(GSEA) which take the expression data to measure the activation of specific pathways [11].

GSEA first will rank all the genes in the expression data by their correlation with the selected phenotype to obtain a ranked gene list. Then the set of genes that belong to the user defined pathways, or pathways defined by public bioinformatics database were selected out in the ranked gene list. Finally, the enrichment score is calculated by walking through the ranked gene list. While walking through the list, when a gene that belongs to the pathway of interest was encountered, the enrichment score (ES) will goes up, or shift away from zero (according to the gene correlation with phenotype, the ES can also goes down). While the gene encountered does not belong to the pathway of interest, the score will then goes down, or shift closer to zero. After getting this score, the following statistic test value was obtained by permuting the gene label within the ranked gene list, according to the null hypothesis that those genes are randomly distributed in the list rather than congregation according to the correlation with phenotype [11].

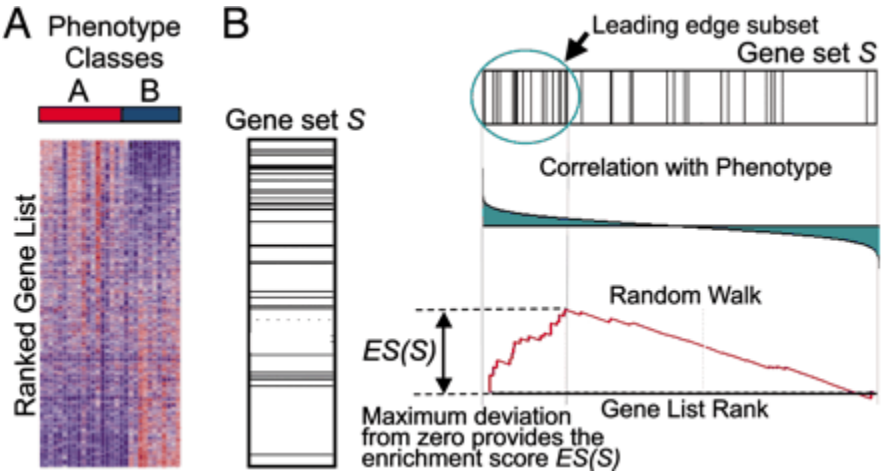


Figure 4. *Subramanian et al*. the GSEA method overview. (A). The heat map of the ranked gene list according to their correlation of phenotype. (B). In the left is the position of each gene in gene set S in the ranked gene list. In the right the walk through the ranked gene list and the red line is the representation of the enrichment score (ES)

2.4 Spearman Correlation

The spearman correlation between calculated SCNA score and the selected ER stress genes were calculated to test if there exists a nice correlation between SCNA and ER stress pathways. Since the SCNA score and the gene expression are not supposed to be within the same numerical scale, a spearman correlation was chosen instead of pearson correlation to obtain a better statistical confidence. The spearman correlation was calculated by the following formula:

$$spearman\ correlation\ x = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2 (y_i - \overline{y})^2}}$$

2.5 Wilcoxon Rank-sum test

The Wilcoxon rank-sum test is applied to test the downstream pathway target of UPR branches. The downstream of UPR branch genes are collected from REACTOME pathway library. Within each tumor types, samples are split into high aneuploidy group and low aneuploidy group. Wilcoxon rank-sum test is applied to test the downstream genes of UPR activation status when comparing their expressions between high aneuploidy group and low aneuploidy group. Since the sample size are limited and not following a suitable hypothesis of t test, the Wilcoxon rank-sum test is applied using the following formula:

The Wilcoxon rank-sum test ranks the all the samples from two groups together, with a null hypothesis that the rank should be randomly distributed rather than biased if they are coming from the same distribution. Then the rank of test group (in this case is high aneuploidy group) was summed together, as the Wilcoxon rank-sum test score. This score would falling into a specific distribution which fit into a table with corresponding p values. From that we could find the p value to validate whether the two groups are coming from the same distribution.

2.6 Network analysis

After the correlation analysis, we made a step further, taking a more systematic view of the ER stress structure, studied the network ER stress under the impact of aneuploidy. In this study, we selected genes in 3 branches of UPR pathway from the REACTOME pathway library, split each cancer types into high and low aneuploidy group based on the previous threshold we discussed (low aneuploidy group which have a SCNA score < 30% of the sample, and high aneuploidy group that have a SCNA

score > 70% of the sample). Then for those selected genes, we conducted a differential co-expression analysis [12].

First, the same method was applied in calculating the spearman correlation, but this time within gene pairs. The spearman correlation between gene pairs x_i and y_j are calculated and stored in a matrix with row number i and column number j:

$$C_{ij} = \frac{\sum_i (x_i - x_{mean})(y_j - y_{mean})}{\sqrt{\sum_i (x_i - x_{mean})^2 \sum_i (y_j - y_{mean})^2}}$$

Then we use the stored correlation matrix, to calculate the matrix of adjacency difference by the formula:

$$d_{ij} = \left(\sqrt{\frac{1}{2} \times \left| sign\left(C^{high}{}_{ij}\right) \times \left(C^{high}{}_{ij}\right)^2 - sign\left(C^{low}{}_{ij}\right) \times \left(C^{low}{}_{ij}\right)^2 \right|} \right)^{\beta}$$

By applying this formula, differences in co-expression pattern between high / low aneuploidy group are been calculated. The higher value of d_{ij} represents a bigger differences of co-expression of gene pairs between high and low aneuploidy groups. Since the value of d_{ij} is between 0 and 1, the bigger beta means a smaller d_{ij} value for small changes in co-expression pattern. Thus, the beta in the formula can be any positive integer that serves as a parameter of soft threshold. After we get the matrix of adjacency difference d_{ij}, a topological overlap matrix T_{ij} for each gene is calculated by the formula:

$$T_{ij} = 1 - \frac{\sum_k (d_{ik} d_{kj}) + d_{ij}}{\sqrt{min(\sum_k d_{ik}, \sum_k d_{jk}) + 1 - d_{ij}}}$$

This matrix T represents the co-expression change of $gene_i$ and $gene_j$ in between the two selected phenotype, high aneuploidy group and low aneuploidy. For example, a small value of T_{ij} means a big change between $gene_i$ and $gene_j$ between two phenotype. One benefit of using the topological overlap matrix is that, it can directly identify genes that share the same neighbors by apply a linkage cluster method on the T_{ij} matrix.

After quantifying the matrix of adjacency differences for each cancer types, we did permutation test to assess the statistical significance of our results. The permutation test is done by generating a 1000 permuted index to randomly select samples, assuming they are from the same group that shares no differences in co-expression pattern. Then we calculate the pairwise module to module co-expression analysis like we did to the high and low aneuploidy groups to see whether the value we get from our previous analysis is significant enough for the following distribution of our permutation test (ie, p value). The importance of statistical significance test is that in the matrix of adjacency, we have a user-defined parameter beta, which is a soft threshold for the differential co-expression analysis that could impact the value of co-expression change. Therefore we should carefully examine the outcome of the test since we induced a beta that is user-defined threshold.

2.7 Immune Phenotype, Linear Model and Partial Correlation

The immune phenotype was analyzed according to *Rooney et al*. [13]. In this paper, the cytolytic activity score is defined to represent the magnitude of immune response within a cell. Since infiltrated immune cells could possibility in a suppressed states which remains in active, cytolytic activity is a better representation of immune response in tumors than directly quantifying the infiltration fractions of immune cells. In this thesis we also choose GZMA and PRF1 as two genes that could represent the immune activity within tumor cells. These two genes are selected out because they showed the highest correlation in cytolytic analysis (nearly 0.9) and also a fold change higher than 10 times when comparing NK cells / cytotoxic T cells with non-hematopoietic tissue.

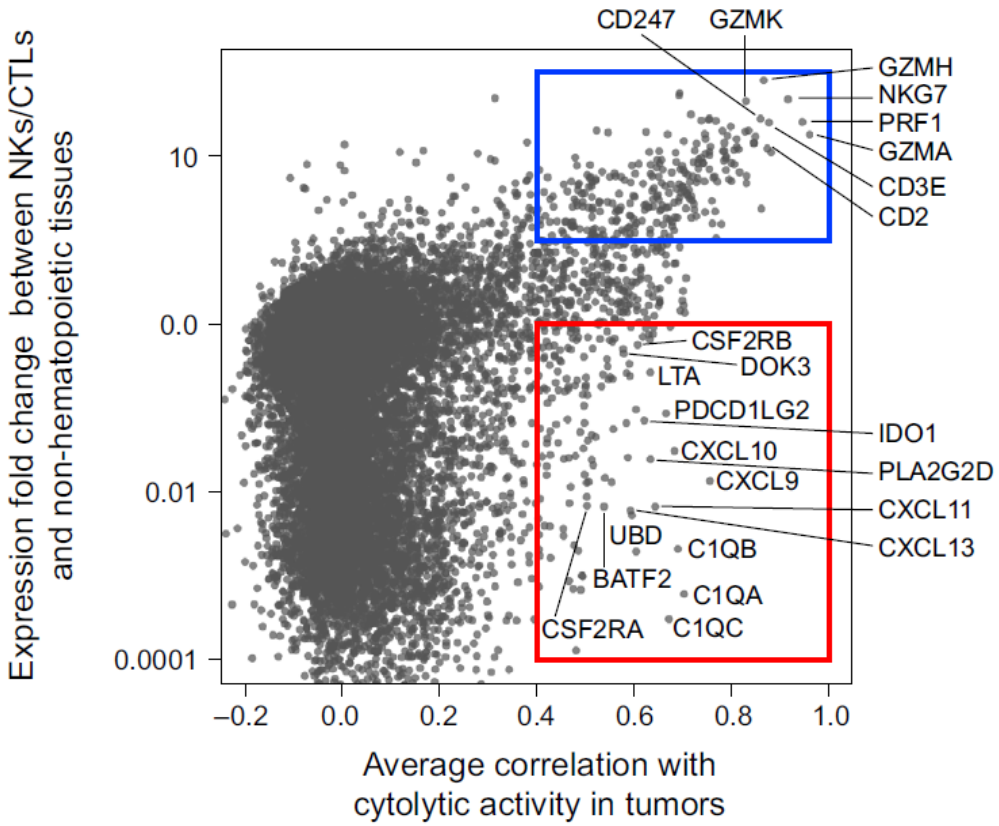


Figure 5. *Rooney et al 2015*. The selection of GZMA and PRF1 as marker of cytolytic activity. X axis is the spearman correlation between genes with cytolytic activity. Y axis is the expression fold change in comparison between NK cells / Cytotoxic T cells and non-hematopoietic tissues. GZMA and PRF1 are been chosen because they have the highest correlation with cytolytic activities in tumors and a fold change over 10.

In general, for every different cancer types the expression of GZMA and PRF1 are been re-scaled within 0 to 10, and then sum up to generate cytolytic activity (CYT) score. Thus, for every i sample, the CYT score is calculated by the following formula:

$$CYT_i = 10\frac{(GMZA_i - GMZA_{min})}{(GMZA_{max} - GMZA_{min})} + 10\frac{(PRF1_i - PRF1_{min})}{(PRF1_{max} - PRF1_{min})}$$

Then the linear regression model was applied to test the relationship between the calculated CYT score and ER stress gene. The linear model formula:

$$CYT = \beta_0 + \beta_i \cdot tumor\ type_i + \sum_j \beta_j \cdot Gene_j \quad , \quad Gene_i \in ER\ stress\ gene$$

Thus, the CYT score was regressed by the sum of an intercept β_0 , the $\beta_i \cdot tumor\ type_i$, and the sum of corresponding ER stress gene multiplied by their coefficient, $\sum_j \beta_j \cdot Gene_j$. Using this formula, applying the least square penalty method, the β_i could represents the CYT score difference between each disease, and the β_j could represents the contribution of each ER stress gene to the SCNA score. We are more interested in the β_j coefficient since we are trying to evaluate the contributions of ER stress in disturbing the immune response. The including of tumor type as a regression factor could diminish the unexplained variance in this model, thus, more confidence on the results of each gene's contribution.

The partial correlation test was applied to test the correlation between SCNA and CYT while setting the ER stress genes expression level in control. With the application of partial correlation, we could understand how much SCNA and ER stress gene independently contribute to the CYT scores while there exists interactions between them. When setting C as the control variable, the calculation of partial correlation between A and B is calculated by:

$$partial\ correlation = \frac{r_{AB} - r_{AC} \cdot r_{BC}}{\sqrt{(1 - r_{AC}^2)(1 - r_{BC}^2)}}$$

Chapter 3 Results

3.1 Aneuploidy, unique signature of cancer, the progression marker, and the key player in impaired immune response

After quantifying aneuploidy, we found that for individual cancer types there exists big differences in different kinds of SCNA level. For instances, kidney chromophobe (KICH), which have the highest ranked arm level SCNA, holding an median value over 31, meaning most of their samples have 31 chromosome arms maintaining an abnormal copy number, have a median value for focal events close to zero. Also within the arm SCNA, thyroid carcinoma and thymoma only have a median value close to 0. By witnessing such a big differences among each cancer types, we treat every cancer types differently in subsequent study in order to avoid a tissue effect differences in SCNA which could cause potential bias in the analysis.

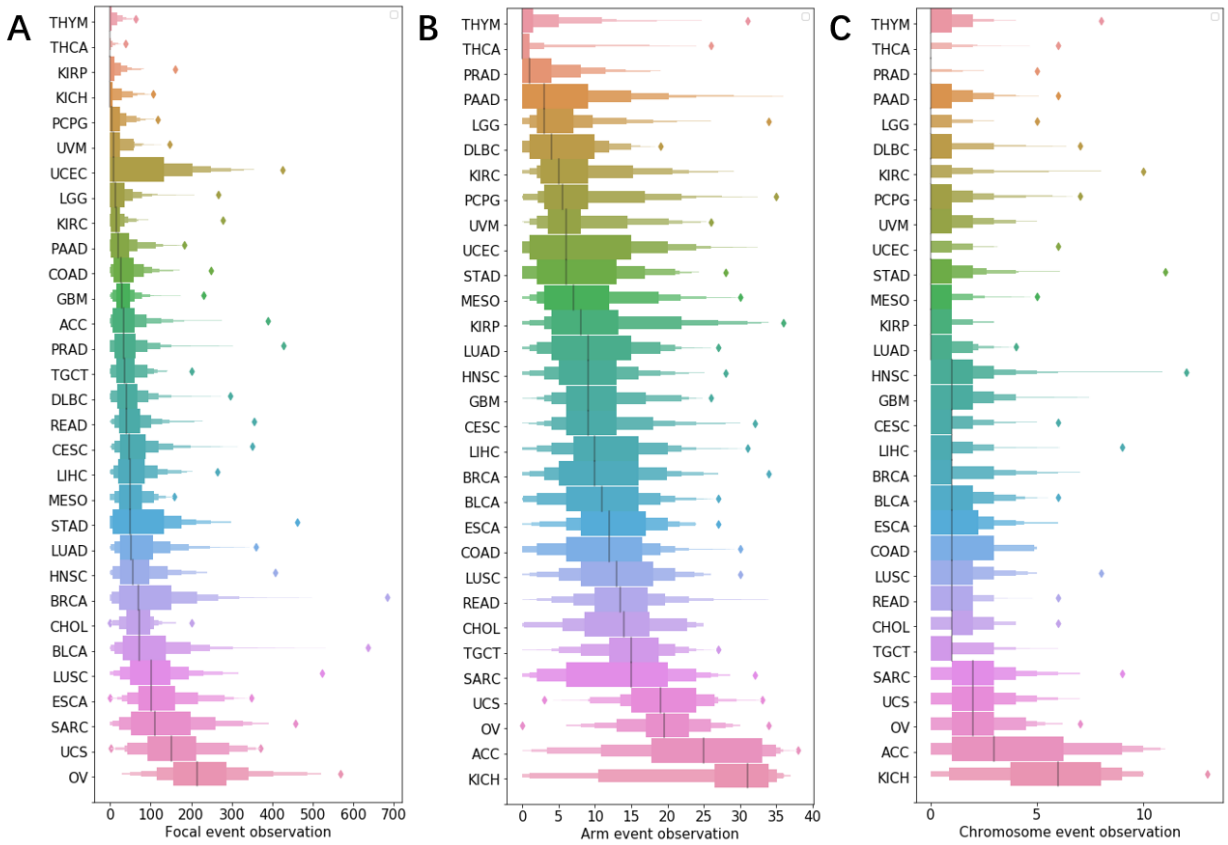
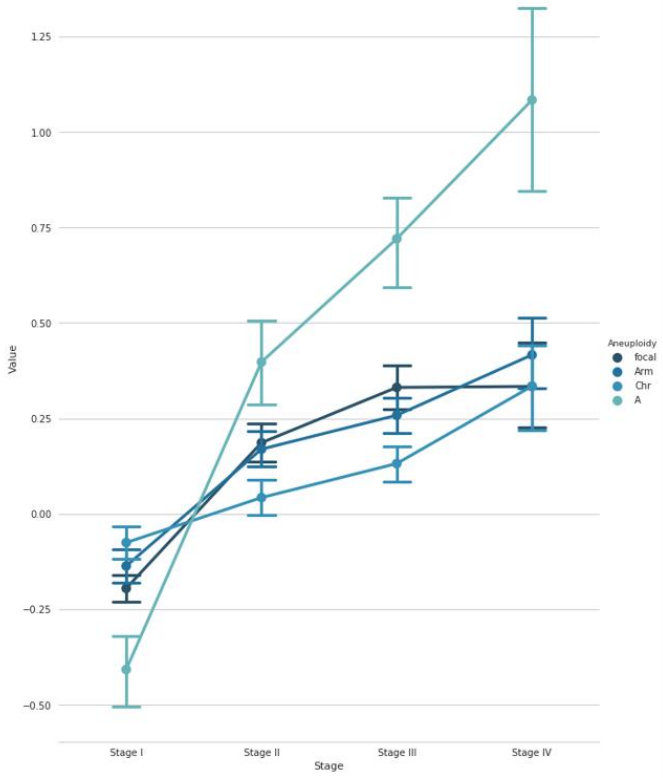


Figure 6. Summary plots showing focal, arm and chromosome level SCNA across different cancer types. The y-axis are cancer types that ranked by their median values. The x axis is the number event that been observed. (A). The focal event ranked by median value from low to high. (B). The Arm level aneuploidy event. (C). The chromosome level aneuploidy event.

In the progression study, we included 5999 samples that with known stage status and tested if aneuploidy could be a progression markers in cancer. We confirm that aneuploidy increases with stages during cancer progression in pan-cancer analysis (Figure 7A). Knowing that aneuploidy level varies across tumor types, we applied the linear regression mode setting the $SCNA$ value as independent variable, and the variable stages and disease as dependent variable to see if we could still observe the stage progression phenomenon while control the tumor types as covariate. In the linear model, most of the disease are showing a p value close to zero, which supported our

previous finding that the $SCNA$ values are different among each cancer types. From stage II to stage IV while setting stage I as control, we show an increasing coefficient and with significant p value (p close to 0) (Figure 7B), meaning by controlling the tumor type differences, the $SCNA$ value showed a developing trend across stages within each tumor types.

A



B

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.8843	0.238	12.105	0.000	2.417	3.351
stage[T.Stage II]	0.4446	0.071	6.265	0.000	0.305	0.584
stage[T.Stage III]	0.4556	0.071	6.379	0.000	0.316	0.596
stage[T.Stage IV]	0.7333	0.093	7.854	0.000	0.550	0.916
disease[T.BLCA]	-3.0890	0.257	-12.025	0.000	-3.593	-2.585
disease[T.BRCA]	-2.9076	0.241	-12.082	0.000	-3.379	-2.436
disease[T.CESC]	-3.3404	0.287	-11.651	0.000	-3.902	-2.778
disease[T.CHOL]	-2.2961	0.403	-5.698	0.000	-3.086	-1.506
disease[T.COAD]	-3.2938	0.252	-13.052	0.000	-3.789	-2.799
disease[T.ESCA]	-2.4506	0.310	-7.911	0.000	-3.058	-1.843
disease[T.HNSC]	-4.0275	0.255	-15.791	0.000	-4.528	-3.528
disease[T.KICH]	1.3417	0.336	3.991	0.000	0.683	2.001
disease[T.KIRC]	-4.3459	0.249	-17.460	0.000	-4.834	-3.858
disease[T.KIRP]	-3.5603	0.267	-13.348	0.000	-4.083	-3.037
disease[T.LIHC]	-3.0721	0.258	-11.909	0.000	-3.578	-2.566
disease[T.LUAD]	-3.2573	0.253	-12.899	0.000	-3.752	-2.762
disease[T.LUSC]	-2.3739	0.254	-9.356	0.000	-2.871	-1.877
disease[T.MESO]	-3.8099	0.318	-11.992	0.000	-4.433	-3.187
disease[T.OV]	0.3467	0.263	1.319	0.187	-0.168	0.862
disease[T.PAAD]	-4.5890	0.278	-16.516	0.000	-5.134	-4.044
disease[T.READ]	-2.5929	0.285	-9.112	0.000	-3.151	-2.035
disease[T.STAD]	-3.6310	0.266	-13.652	0.000	-4.152	-3.110
disease[T.TGCT]	-2.2486	0.321	-7.009	0.000	-2.877	-1.620
disease[T.THCA]	-5.6667	0.251	-22.601	0.000	-6.158	-5.175
disease[T.UCEC]	-3.1844	0.257	-12.403	0.000	-3.688	-2.681
disease[T.UCS]	-0.4167	0.383	-1.088	0.276	-1.167	0.334
disease[T.UVM]	-4.5846	0.320	-14.313	0.000	-5.213	-3.957

Figure 7. The aneuploidy as a progression marker in pan-cancer study. (A). The x axis are from stage I to stage IV categories. The y axis are the$A$ values we calculated (the A_i value was calculated according to method 2.1 mentioned above). The color from dark to light represents focal aneuploidy level, arm aneuploidy level, chromosome aneuploidy level and the A_i value. (B). The summary table for the linear regression model for aneuploidy. Coefficient, standard error, t statistical value, p value and confidence interval are included in each columns. The corresponding rows are the variables tested in the linear model, see method 2.1.

After quantifying focal, arm and chromosome level SCNA for individual patients, we generated the arm level SCNA signature for each cancer types by clustering the arm

level SCNA within cancer types. Our studies is highly similar to *Taylor et al* who applied a mixture gaussian model to generate the signature for different cancer types [10]. The signal of chromosome 3q copy number gain in lung squamous cell carcinoma (LUSC) is strong in both analysis.

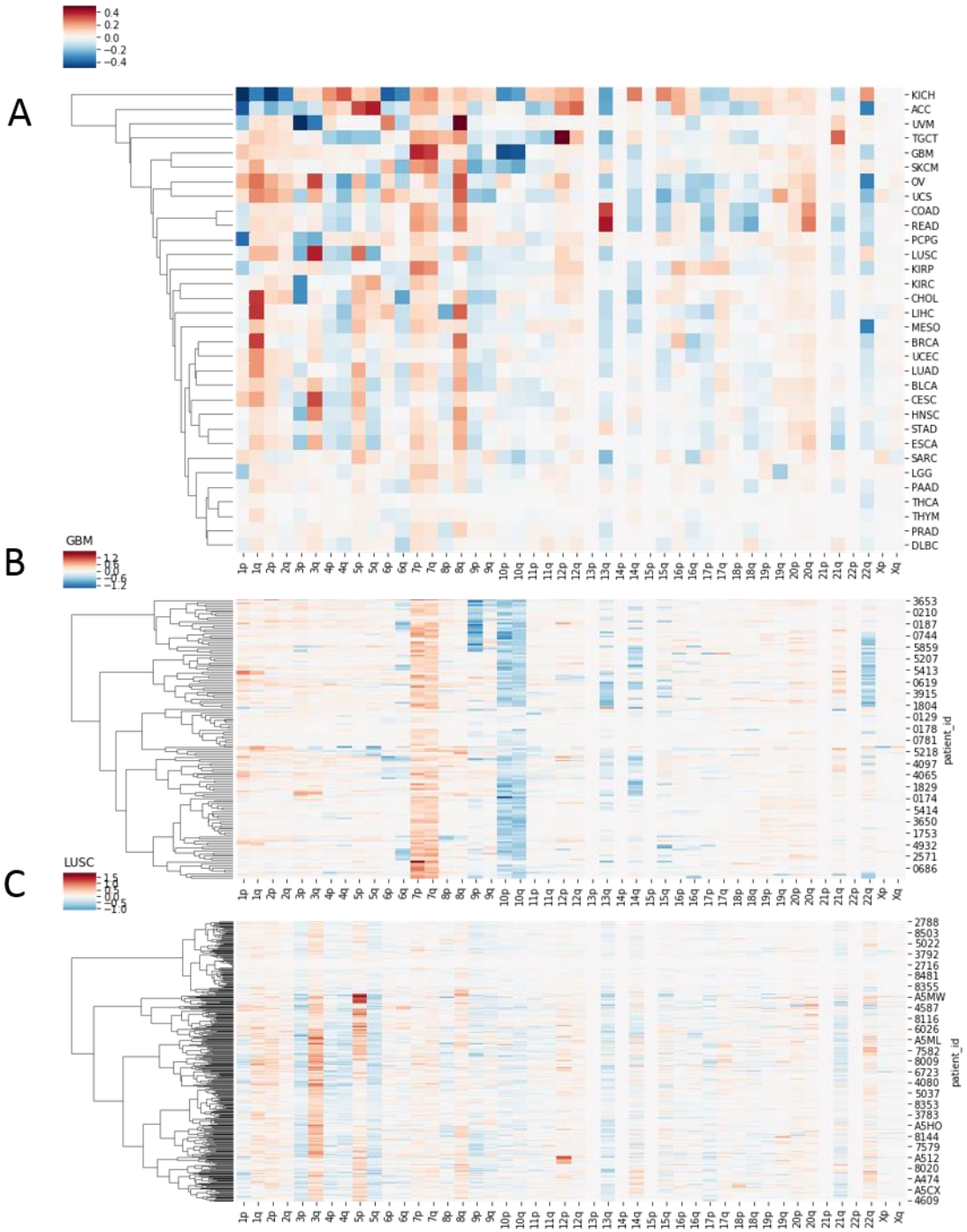


Figure 8. Arm level signatures. From top to bottom, the total signature across All 31 cancer types (A); the arm level signature for GBM (B) and for LUSC (C). The x axis represents the index of chromosome from 1p to Xq. The y axis in (A) represents each corresponding cancer types. The y axis in (B) and (C) are the patient id. (A). Total arm SCNA signature of 31 cancer types. Each cell represents the copy number change averaged across patient within this cancer type. (B). The arm level signature for glioblastoma (GBM). A strong signal of chromosome arm copy number gain in 7p,7q

and decrease in 10p 10q are observed in this disease. (C). Strong signal of copy number gain in 3q can be witnessed in lung squamous cell carcinoma (LUSC), those signals might suggest a selection mechanisms in tumor progression.

Then we did a gene set enrichment analysis (GSEA) for every cancer types we have using the expression data. We set our control to be the low aneuploidy group (SCNA < 30% of the sample), and the comparison group to be the high aneuploidy group (SCNA > 70% of the sample). In this analysis, among 31 cancer types, 17 of them are witnessed a significant gene expression decline within antigen presenting pathway in high aneuploidy groups, when compared with low aneuploidy groups, which suggest that higher aneuploidy are affecting the ability of tumor cells to present antigen appropriately. Meanwhile, we also tested several related immune pathways. NK cell cytotoxicity pathway, T cell receptor pathway and B cell receptor pathway are all downregulated in the high aneuploidy group for most of the tumor types. This analysis is quite similar to *Davoli et al* who also noticed a negative correlation between SCNA and immune response [3].

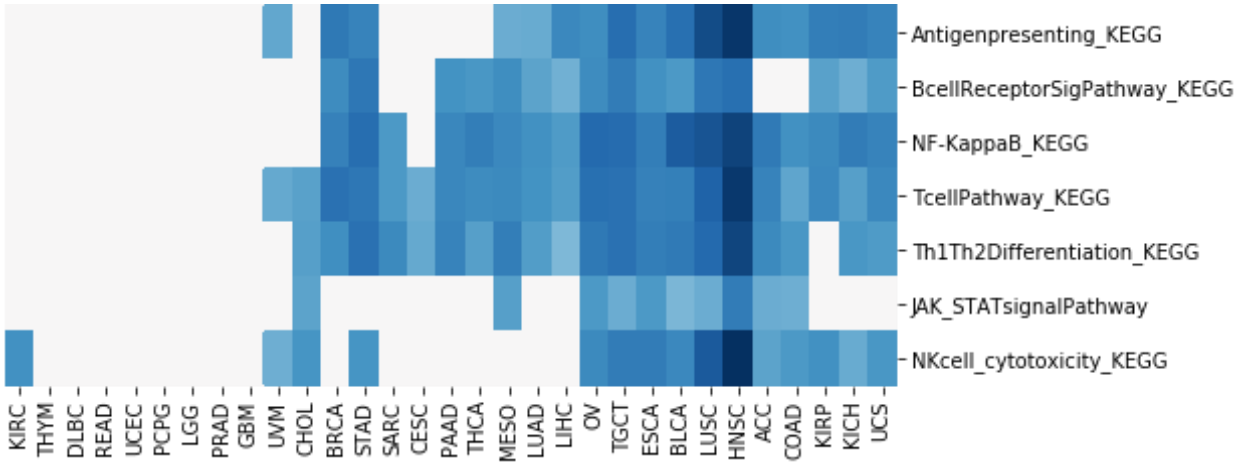


Figure 9. The summary heat map of GSEA results. Each columns stands for a specific cancer type. Each row represents the activation signals of specific pathways. The blue

cells are pathways that significantly downregulated in high SCNA group in comparison with low SCNA group. White cells are insignificant results.

3.2 The activation of ER stress and the its network structure

In order to test whether the ER stress is present in tumor samples, we need to test the expression of ER stress marker genes and compare it to the normal sample level. Since the upregulation of HSPA5 is agreed to be both the sign of ER stress and the activation of UPR, we applied t test to evaluate the expression differences of HSPA5 between tumor and its matched normal samples.

In the t statistic test, HSPA5 is significantly upregulated among 12 tumor types and downregulated in 3 tumor types (THCA, KIRP, KICH). While there are 7 insignificant studies, this could be caused by the limited sample size of corresponding normal tissue. For example, we only have 2 normal samples for SARC and 3 normal samples for PCPG which could drastically decrease the statistic powers in this analysis. In this study, the significant upregulation of HSPA5 across various tumor types is undoubtedly a proof of the presence of ER stress and its activation of rescue pathway UPR.

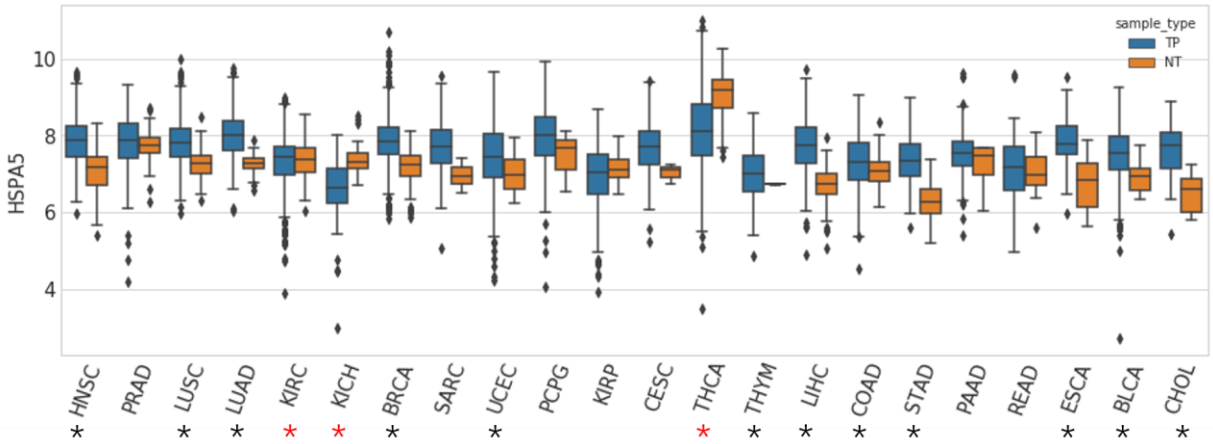


Figure 10. Boxplot of HSPA5 gene expression in 22 tumor types. The value in y axis is the log2 level of HSPA5 expression. X axis correspond to each tumor name. The box in blue is the tumor samples and box in orange are corresponding normal tissues within the same tumor type. Black asterisk suggests a significant increase in HSPA5 gene expression within this tumor samples when compared with its matched normal samples. Red asterisk suggests a significant decrease in HSPA5.

As the presence of ER stress is witnessed, the activation of UPR is witnessed, we continued our analysis testing if the activation of UPR genes are correlated with aneuploidy. Spearman correlation between the selected ER stress genes and the total SCNA scores are calculated within individual cancer types. The heat map suggests that a strong correlation in EIF2α and DDIT3, and some mild correlation with ATF6 between SCNA scores are been found in most cancer types. Both DDIT3 and EIF2α are genes within the PERK downstream pathway. Signals from ERN1 pathway is nearly lost, which might suggest a lacks of activation in ERN1 pathway. It is interesting that aneuploidy favors PERK more than two others in UPR.

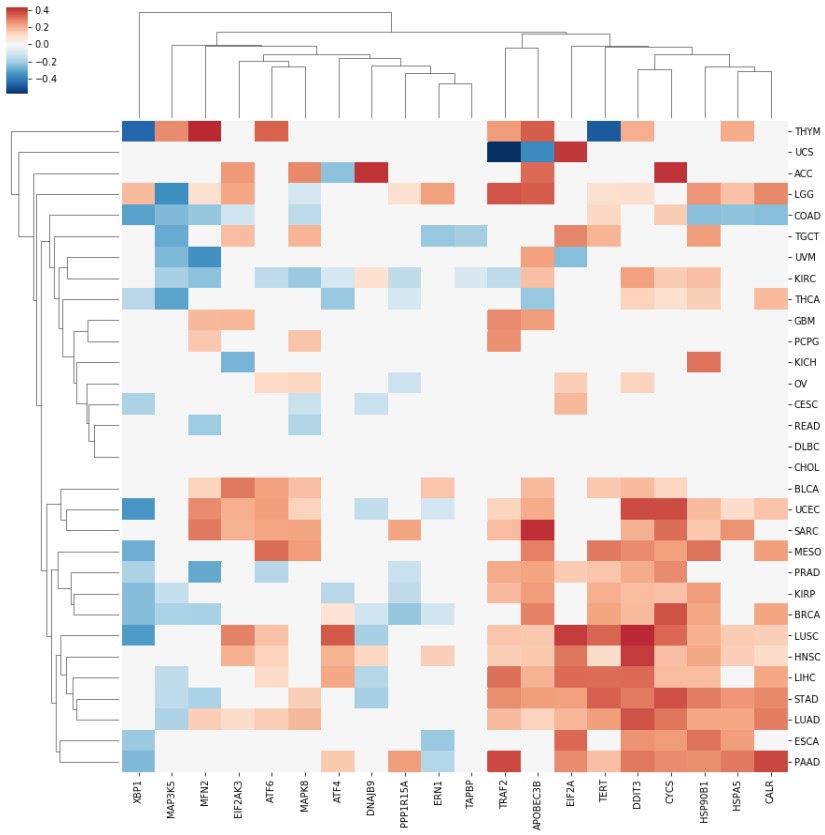


Figure 11. Heat map of the spearman correlation between SCNA and selected ER stress genes within each cancer types. Red cells suggesting the gene (in row) within the corresponding tissue type (in column) is significant up regulated in high aneuploidy group while blue cells suggest a significant decrease. White cells are non-significant study.

In Figure 11 we witnessed a signal lost from ERN1 branch of UPR which is interesting. As ERN1 pathway activation needs post-translational modification, it is also necessary to test the expression level of its downstream target genes. We then collected gene set in the downstream pathway of ERN1 branches from REACTOME and applied Wilcoxon rank-sum test to test the expression differences of those downstream genes between the high aneuploidy group and low aneuploidy groups for each tumor types. In Figure 12, the heat map showed that across half of the cancer

types we saw a large amount of genes are significantly increased in expression level, which is an approval of ERN1 pathway activation.

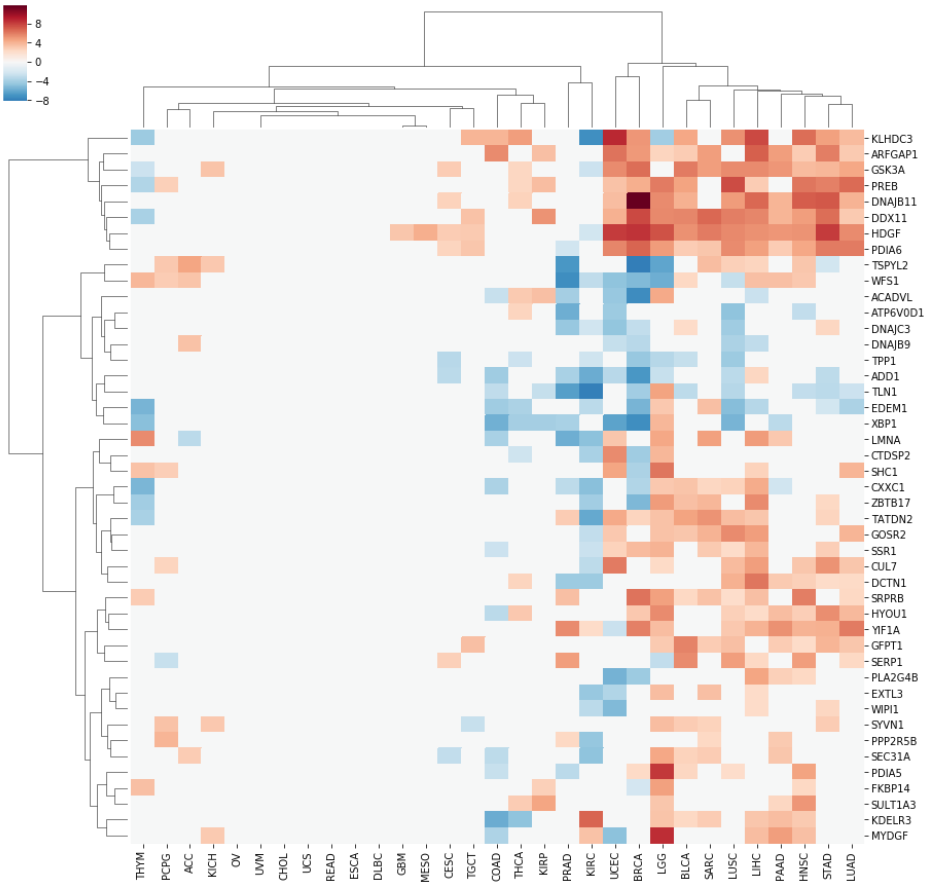


Figure 12. Heat map of Wilcoxon rank-sum test of the expression differences of ERN1 downstream targets genes between high SCNA and low SCNA groups for each tumor types. Red cells suggesting the gene (in row) within the corresponding tissue type (in column) is significant up regulated in high aneuploidy group while blue cells suggest a significant decrease. White cells are non-significant study.

After the expression analysis, we also applied the network study to test if there exists different UPR pathway structures between the high aneuploidy group and the low aneuploidy group within each tissue types. We selected genes in 3 branches of UPR pathway from the REACTOME pathway library, split samples into high and low aneuploidy group based on the previous threshold we discussed (low aneuploidy group which have a SCNA score < 30% of the sample , and high aneuploidy group that have a

SCNA score > 70% of the sample) within individual cancer types. Then for those selected genes, we conducted a differential co-expression analysis. The network study showed a great gene network co expression loss in higher aneuploidy groups across a lot of cancer types, especially in ERN1 and ATF6 pathways.

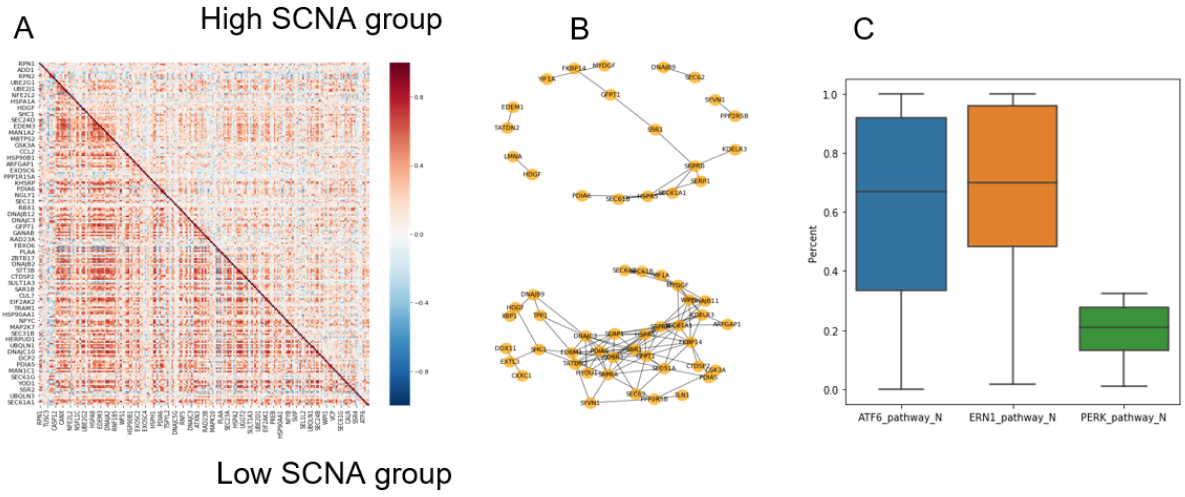


Figure 13. High aneuploidy correlated with perturbed network structure in ERN1 and ATF6 pathways. (A). The co-expression pattern heat map of genes with significant co-expression pattern changes. Upper triangle is the high SCNA group co-expression which is more whiter than the Low SCNA group in the bottom triangle. (B). Network plot of genes in ERN1 pathways in sarcoma (SARC). The top part is the high aneuploidy group that genes are less connected. The bottom part is the low aneuploidy group. (C). Pan-cancer average percentage of genes with modified co-expression patterns in UPR pathways.

3.3 The bridge between aneuploidy and impaired immune response

In section 2.6 we discussed the method applied to measure the immune response which gave us every individual sample a CYT score. Linear model was then

applied, in order to test the contribution of different ER stress genes to the impaired immune response.

Table 1. Coefficient for UPR genes in the applied linear model.

Gene	Coefficient	Standard Error	T test statistics	P value	CI (0.025	CI (0.975
ERN1	-0.0168	0.049	-0.346	0.730	-0.112	0.079
ATF4	-0.0466	0.053	-0.887	0.375	-0.150	0.056
ATF6	-0.2268	0.062	-3.635	0.000	-0.349	-0.104
EIF2AK3	-0.2469	0.052	-4.726	0.000	-0.349	-0.144
XBP1	0.0830	0.030	2.735	0.006	0.024	0.142
EIF2α	-0.4921	0.050	-9.916	0.000	-0.589	-0.395
DDIT3	-0.2342	0.033	-7.042	0.000	-0.299	-0.169
HSPA5	0.1552	0.060	2.588	0.010	0.038	0.273
DNAJB9	0.5039	0.046	10.957	0.000	0.414	0.594
PPP1R15A	0.0931	0.031	3.018	0.003	0.033	0.154
HSP90B1	0.1968	0.058	3.379	0.001	0.083	0.311
TRAF2	0.7486	0.047	15.994	0.000	0.657	0.840
MAP3K5	0.3822	0.035	10.895	0.000	0.313	0.451
MFN2	-0.1824	0.050	-3.614	0.000	-0.281	-0.083
CYCS	-0.1245	0.041	-3.036	0.002	-0.205	-0.044
CALR	-0.0415	0.056	-0.739	0.460	-0.151	0.069

Since we are not sure if the CYT also varies for different tumor types, within the linear model tumor types are also been set as covariate. Several disease in this model showed a significant differences in CYT level, except LGG (p value = 0.747) and PCPG (p value = 0.851), suggesting CYT score do show differences among every tumor type. With the tumor type as covariate in this model, we still found that almost all the ER stress genes are showing significant p values in the linear model, except ERN1, ATF4 and CALR. In this study, PERK, EIF2α, ATF6 and DDIT3 showing a strong negative coefficient which suggest a strong negative correlation with the immune response.

After found EIF2AK3, ATF6, DDIT3 and EIF2α we performed a spearman correlation study to examine the relationships among aneuploidy, selected genes and cytolytic activity. In the spearman correlation study (Figure 14), SCNA scores are negatively correlated with CYTs in 29 tumor types, which also supports our previous findings that high aneuploidy are correlated with decreased antigen presenting pathways (Figure 9). Then, 20 cancer types showed a negative correlation between DDIT3 and CTY, and 27 cancer types showed a negative correlation between EIF2α. While EIF2AK3 and ATF6 are showing more noisy results, tend to be more positively correlated with CYT scores in several tumor types.

We also noticed that when there exists 2 or more variables interacts together to affect the third or another dependent variable, partial correlation should be applied to test how much of the contribution is been made by each independent variables. The partial correlation can be used to calculate the correlation between X and Y when giving a controlling variable Z. Since DDIT3 and EIF2α are correlated with SCNA (Figure 11), and showing a stronge negative correlation with CTY, we set the gene expression as

control variable separately (DDIT3, EIF2α) to test the partial correlation between SCNA and aneuploidy.

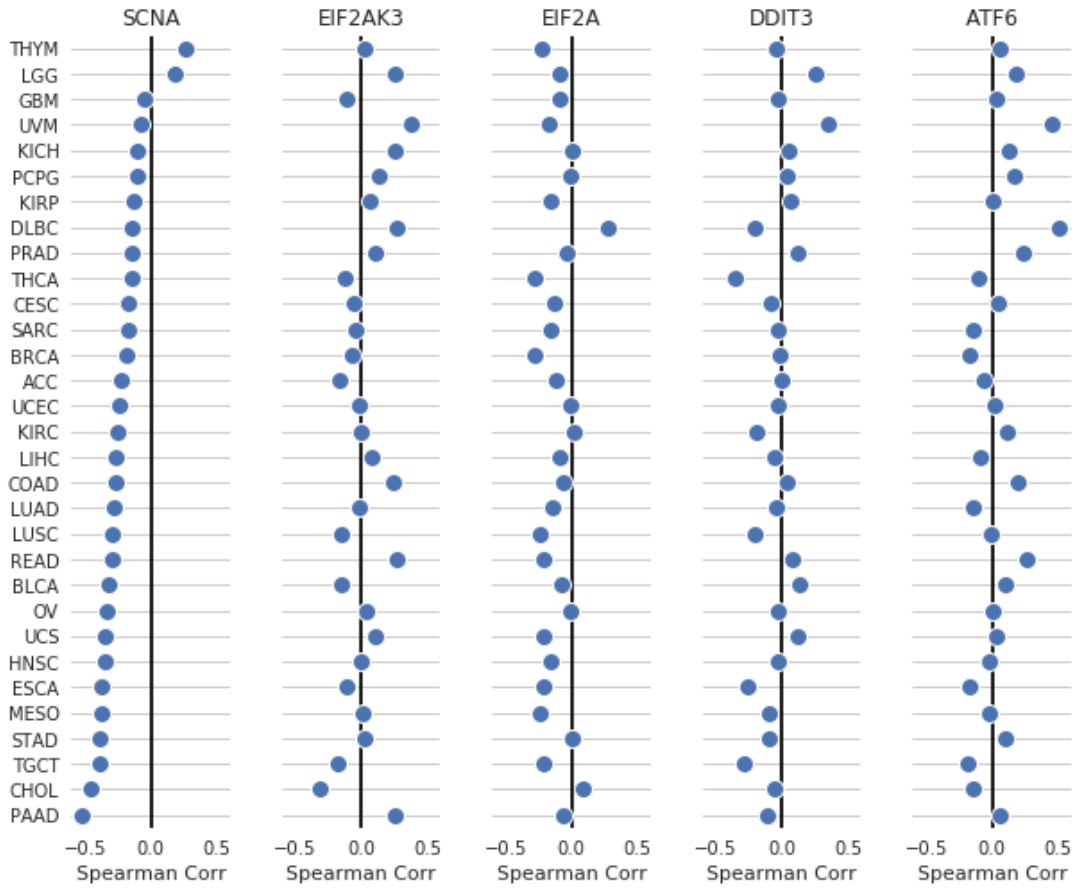


Figure 14. Spearman correlation analysis between SCNA, selected genes and CYT score. SCNA showed negative correlation in 29 cancer types with CYT score. EIF2α showing 27 and DDIT 20.

Figure 15 shows the partial correlation between SCNA and CYT when setting EIF2α or DDIT3 as control variable. DDIT3 and EIF2α are both used as control variable to test the partial correlation between SCNA and CYT since they both show high correlations with SCNA and CYT. In this analysis we notice that for some tumor types like PAAD, CHOL, TGCT, THYM and LGG the partial correlation is very different from spearman correlation. This result suggests an interaction between SCNA and EIF2A or

DDIT3 in affecting CYT score. In THYM and LGG we saw the spearman correlation between SCNA and CYT are positive. While the partial correlation between SCNA and CYT when setting EIF2α and DDIT3 as control shows a negative value, suggesting that SCNA are more negatively correlated with CYT when the interactions with EIF2α and DDIT3 are excluded. Interestingly, for PAAD the spearman correlation is the strongest one across 31 cancer types. While we set EIF2α or DDIT3 as control the partial correlation are close to zero, which suggest EIF2α and DDIT3 are stronger factors than SCNA that contribute to the negative correlation between SCNA and CYT. Thus, for different cancer types the interactions between SCNA and ER stress various. Some tumor types are like THYM and LGG that SCNA contributes more to the negative correlation between CYT. Others are like PAAD, which EIF2α and DDIT3 contributes more to the negative correlation between CYT.

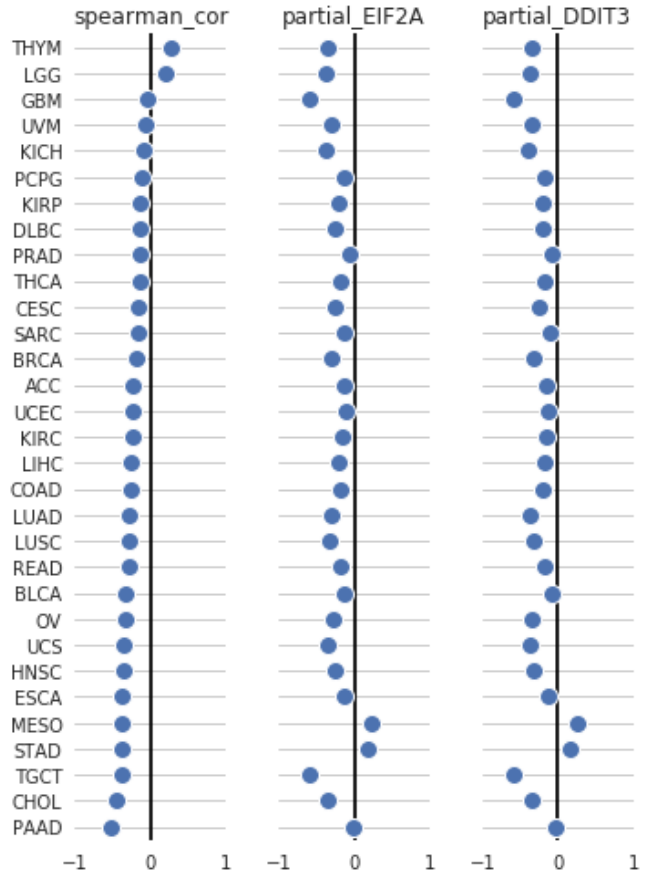


Figure 15. Partial correlation plot. The top label in x axis represents the related pathway that set as control variable in partial correlation study. Y axis represents each tumor type. The left one is the same spearman correlation between SCNA and CYT as in Figure 14.

Chapter 4 Discussion

4.1 Aneuploidy, the puzzle and the future

In the previous chapter (chapter 2.1) we defined our methods for quantifying aneuploidy. The term aneuploidy has been in use for decades. However, the computational study and quantification of aneuploidy from data generated by modern sequencing technologies is still an area of active development. In the context of cancer, the term aneuploidy is frequently used in a broader sense and can include two major phenomena: focal copy number alterations and chromosome arm alterations. These different types of alteration likely result from different mechanisms and contribute to cancer progression in different ways [3]. In our case, we divided aneuploidy into three categories, including focal level SCNA, chromosome arm level SCNA and chromosome SCNA (which is similar to chromosome arm level SCNA), taking both gain and loss of copy number into account, as was described in the study by *Davoli et al*.. The main difference is that *Davoli et al.* applied a tool called ABSOLUTE to quantify sample purity and adjusted the copy number change threshold according to sample purity. However, we and others have encountered difficulties replicating published results from previous studies with the ABSOLUTE package. Therefore instead of calculating purity with ABSOLUTE, we applied a threshold as previously described by *Beroukhim et al* to call copy number alteration events [8].

Despite the differences in the method of quantifying aneuploidy level, we also found evidence that aneuploidy was inversely correlated with measures of effective immunity. These results are similar to those reported by *Davoli et al*, and confirm the immune-suppression phenotype of highly aneuploidy tumors. According to *Davoli et al*,

9 out of 12 cancer types had a significant decrease in the CD8 T cells [3]. Our analysis also revealed reduced levels of CD 8 T-cells in tumors with higher levels of aneuploidy. This suggests that differences in calling SCNA events are not large enough to confound the underlying biological relationships.

Interestingly, in a recent publication in 2018, *Taylor et al* applied a gaussian mixture model to quantify chromosome arm level SCNAs based on their definition of aneuploidy. But instead of defining a threshold like us or *Davoli et al*, Taylor et al. used the total ploidy level calculated from ABSOLUTE as a reference value of total tumor genome ploidy to quantify arm level SCNA. Once again, despite different methods for quantifying chromosome arm SCNA level, we reached a highly similar ranking of the chromosome arm SCNA levels in the 31 cancer types used in this analysis (Table 2). Moreover, their arm level signature study is pretty similar to ours, with both studies identifying the 3q and 3p signature in LUSC [10].

Table 2. Summary table of the comparison of total arm level SCNA ranking between Taylor et al. and our study. In this table most of the ranking are fairly close, suggesting that the two different methods both captured the feature of aneuploidy. Red and marked with * are tumor types that have a larger difference in ranking.

Tumor type	Ranking in Taylor et al.	Ranking in our study
TGCT	1	6
ACC	2	2
UCS	3	4
LUSC	4	9
LUAD*	5	18
KICH	6	1
ESCA	7	11

Table 3. Summary table of the comparison of total arm level SCNA ranking between Taylor et al. continued.

READ	8	8
OV	9	3
BLCA	10	12
SKCM	11	na
BRCA	12	13
COAD	13	10
HNSC	14	17
STAD	15	21
SARC	16	5
LIHC	17	14
CHOL*	18	7
CESC	19	15
PAAD*	20	28
KIRP	21	19
GBM	22	16
MESO	23	20
KIRC	24	25
DLBC	25	26
UCEC	26	22
PCPG	27	24
UVM	28	23
LGG	29	27
THYM	30	31
PRAD	31	29
LAML	32	na
THCA	33	30

Thus, there exist various methods to quantify aneuploidy level that return similar results when applied to TCGA. The limitation in developing a perfect model is mainly caused by the need to define a reasonable threshold for calling a copy number change in this dataset with no ground truth to validate against. However, given the relative agreement between methods, we believe our estimates of aneuploidy level are sufficiently reliable for the analyses we performed.

4.2 Aneuploidy and the immune response

In order to assess the relationship between aneuploidy and immunity, we first applied GSEA analysis to test the differences in immune responses between low and high aneuploidy tumors for each cancer type. We focused only on pathways associated with immunity, reducing the total burden of multiple testing correction. We found that 17 out of 31 cancer types had a significant decrease in expression of genes participating in the antigen-presenting pathway, which is in accordance with previous studies [3].

Next we identified other measures of immune activity that could be derived from RNA sequencing data. The first measure, the CYT score, is estimates cytolytic activity from the expression of two genes, PRF1 and GZMA [13]. In this analysis, 29 out of 31 cancer types showed a negative correlation between the SCNA score and CYT score, indicating less CD8 T-cell driven cytotoxicity in the presence of higher levels of aneuploidy, and supporting that aneuploidy is decreases immune response. We also measured immune cell infiltration based on expression levels of immune cell type-specific cell surface markers, and observed that multiple immune cell populations were present at reduced levels in high aneuploidy tumors in multiple tumor types. Thus our

analyses were in general agreement, supporting reduced antigen presentation, reduced levels of immune cells and reduced cytotoxicity in the presence of aneuploidy.

4.2 The bridge between aneuploidy and immune response: the hidden network of ER stress

In the search for mechanisms by which aneuploidy could lead to a decrease in immune response, we focused on genes that were implicated in previous work as playing important roles in the ER stress response. We found that HSPA5 is up-regulated in tumor samples relative to corresponding tissue-matched control samples, a sign of UPR activation. We also observed that the expression of several ER stress pathway genes, EIF2AK3, EIF2α, ATF6 and DDIT3, was strongly correlated with SCNA scores across various cancer types. However, counter to our expectation, we did not observe a correlation between expression of ERN1 pathway genes and SCNA levels. Because the ERN1 pathway is activated by post-translational modifiers, we speculated that the expression of genes regulated downstream of the ERN1 pathway might provide a better indication of ERN1 pathway activation status. We found such a gene set in the REACTOME library (R-HSA-381038) called XBP1s activated chaperone genes [14]. Analysis of downstream target gene expression revealed a significant increase, suggesting that this pathway is also activated in at least half of the tumor types. Thus, RNA-seq analysis of ER stress genes themselves is not sufficient to determine ERN1 pathway activity, since the mechanism of activation is through the phosphorylation of ERN1 and the alternative splicing of XBP1 to XBP1s, rather than up-regulation. In this way, we confirmed that all 3 branches of the ER stress response are activated and correlate with the SCNA scores in tumors.

Apart from expression and pathway analysis, we also did a co-expression analysis to evaluate the direct effects of aneuploidy on the ER stress pathway. Biological systems require balanced stoichiometries to effectively transmit information. Thus genes within the same pathway should be co-expressed when the pathway is active. Aneuploidy can cause imbalances in stoichiometry by altering the copy numbers of genes encoding protein components of the pathways. We thus compared co-expression patterns among ER stress genes among high aneuploidy tumors versus low aneuploidy tumors for each tumor type. This analysis revealed a large reduction in the amount of gene co-expression in high aneuploidy tumors relative to low aneuploidy tumors. Interestingly, when the three ER stress branches were considered separately, we observed substantially more co-expression preserved in the PERK pathway as compared with the ERN1 and ATF6 pathways in various cancer types. This raises the question of whether preserving signal transduction through the PERK arm of ER stress confers an advantage to tumors. The phosphorylation of EIF2α within the PERK pathway is believed to play a pivotal role in solid tumor growth, invasion and angiogenesis. There is also evidence that EIF2α phosphorylation is a master regulator of cell adaptation to ER stress conditions [15]. In contrast, DDIT3 - also within PERK pathway - was suggested to initiate apoptosis under pro-longed ER stress state, and can regulate other cellular functions that are not relevant to apoptosis [16]. In our study the up-regulation of DDIT3 did not correlate with an apoptotic signal. These results suggest that PERK may play an important role in tumor survival and progression, and as such this pathway may merit investigation as a target for anti-cancer therapies.

After establishing the link between aneuploidy and ER stress responses, we next evaluated the relationship between ER stress and measures of immune response. Since cytolytic activity is likely a more representative measure of anti-tumor immune activity than immune infiltration alone, we used the CYT score as the metric of immune response. We applied linear models to study the relationship and interactions among ER stress genes and the CYT score. The expression of several ER stress genes was negatively correlated with CYT, with the strongest correlations being observed for DDIT3 and EIF2α. This suggested that the PERK pathway may be the predominant pathway that contributes to an impaired immune response downstream of ER stress induced by aneuploidy.

4.3 Bringing it all together

This project began with the hypothesis that ER stress is an intermediary between aneuploidy and impaired immune responses based on findings by Zanetti et al that supported a link between ER stress signaling and immune suppressive effects. We confirmed that the SCNA scores were negatively correlated with CYT scores in most of the cancer types and positively correlated with ER stress activities. DDIT3, EIF2AK3, ATF6 and EIF2α were highly correlated with SCNA scores across 22 cancer types. We also observed that several ER stress genes were inversely correlated with cytotoxic activity, with DDIT3 and EIF2α showing the strongest effects. This led us to focus on the PERK brank of ER stress as the likely intermediary between aneuploidy and immune response.

DDIT3 and EIF2α are both regulatory transcription factors [15, 16]. The increase of DDIT3, witnessed in numerous cancer types, is a sign of long term unrelieved ER stress suggesting a chronic stress state of cells. In addition, the phosphorylation of EIF2α is shown to be necessary for the growth of solid tumors [7]. We observed increasing aneuploidy with tumor stage, and a positive association between DDIT3, EIF2α and SCNA scores. Together these observations suggest that increasing levels of aneuploidy lead progressively to chronic stress during tumor progression which is reflected by PERK activity.

In order to evaluate the possibility that DDIT3 and EIF2α levels mediate reduced immunity in the presence of aneuploidy, we performed a partial correlation test. The observed reduction in correlation between SCNA and CYT score when conditioning on expression of DDIT3 and EIF2α supports the possibility that the PERK branch of the ER stress pathway provides a mechanism by aneuploidy can result in the suppression of anti-tumor immunity. This result sets the stage for follow on experimental validation studies.

Interestingly, by measure of gene co-expression the PERK pathway was the most intact compared with disturbed ERN1 and ATF6 pathways in high aneuploidy tumors. Since PERK is universally activated in various cancer types, it is important to understand how cancer cells take advantage of this pathway to promote growth and escape immune response. Although the ERN1 and ATF6 pathways were frequently perturbed in high aneuploidy tumors, we still noticed that these pathways showed signs of increased activity and correlation with aneuploidy in tumors. Therefore, more careful

analysis of these branches of ER stress will be necessary to rule out their relevance to immune response.

There are some limitations of this work that must be taken into consideration. For statistical power, we grouped samples according to tumor type, however this ignores that there are distinct subtypes that could have different characteristics at the level of aneuploidy, ER stress or immune response. Within this analysis, signal loss in the correlation studies suggests, a step by step subtype analysis within tumor type may be necessary to cleanly dissect the relationship between ER stress and aneuploidy. Thus, this study provides a preliminary concept in bulk tumor that provides a strong basis for more careful follow on studies to determine the precise mechanisms by which ER stress mediates aneuploidy-driven immune suppression.

Chapter 5.Conclusion

In this thesis, firstly, we show that tumor aneuploidy is correlated with stage which suggest that aneuploidy could serve as a marker of tumor progression in many tumor types. Then ER stress was found to be correlated with aneuploidy. The 3 branches of ER stress were all activated in tumors and showed correlation with aneuploidy, in which DDIT3 and EIF2α in the most preserved PERK pathway interact with aneuploidy to affect the immune responses across several cancer types. This validates our model, supporting that ER stress plays role in connecting aneuploidy with impaired immune response. However, while confirming the bridge between aneuploidy and immune response, we should still be open to other possible mechanisms by which aneuploidy leads to decreased immune activity during tumor progression. Because biological systems are a combination of numerable complex regulatory networks, there could also exists several hidden structures under this bridge, or co-existing parallel connections in the midst of aneuploidy and immune response, and ER stress is only one of them.

Briefly, in this thesis we discussed the ER stress as a possible mechanisms that could connect aneuploidy to impaired immune response. Strong evidence suggesting that within the PERK pathway, DDIT3 and EIF2α are up regulated and interact with aneuploidy in tumor cells. Those two genes interact with aneuploidy and with each other, showing a significant negative correlation across various tumor types. Interestingly, apart from the PERK pathway, the structure of other two pathways, ERN1 and ATF6, are both modified by aneuploidy. A large amount of gene-gene co-expression pattern loss in high aneuploidy group in these pathways was evident in

numerous tumor types. We speculate that such an analysis could be applied more broadly to reveal other pathways and parts of pathways that are preserved despite the increasing genomic chaos that results from progressive aneuploidy.

Though identified up-regulation of DDIT3 and EIF2α across various tumor types and their interactions with aneuploidy as affecting the immune response, we did not fully unravel the mechanisms by which DDIT3 and EIF2α modify the immune response of cancer cells during tumorigenesis. As DDIT3 and EIF2α are both transcriptional regulators, future studies could focus on the role of their downstream targets in tumor growth. The network interaction loss of the ERN1 and ATF6 pathways may also merit careful exploration since the up-regulation of these two pathways are also observed, and the network loss of particular interactions within those pathways might also provide selective advantages in tumor progression.

Bibliography

1. Santaguida, S., & Amon, A. (2015). Short-and long-term effects of chromosome mis-segregation and aneuploidy. *Nature reviews Molecular cell biology*, *16*(8), 473.

2. Janssen, A., van der Burg, M., Szuhai, K., Kops, G. J., & Medema, R. H. (2011). Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science*, *333*(6051), 1895-1898.

3. Davoli, T., Uno, H., Wooten, E. C., & Elledge, S. J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science*, *355*(6322), eaaf8399.

4. Zhu, J., Tsai, H. J., Gordon, M. R., & Li, R. (2018). Cellular Stress Associated with Aneuploidy. *Developmental cell*, *44*(4), 420-431.

5. Zanetti, M., Rodvold, J. J., & Mahadevan, N. R. (2016). The evolving paradigm of cell-nonautonomous UPR-based regulation of immunity by cancer cells. *Oncogene*, *35*(3), 269.

6. Granados, D. P., Tanguay, P. L., Hardy, M. P., Caron, É., de Verteuil, D., Meloche, S., & Perreault, C. (2009). ER stress affects processing of MHC class I-associated peptides. BMC immunology, 10(1), 10.

7. Yoo, Y. S., Han, H. G., & Jeon, Y. J. (2017). Unfolded Protein Response of the Endoplasmic Reticulum in Tumor Progression and Immunogenicity. *Oxidative medicine and cellular longevity*, *2017*.

8. Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., ... & Mc Henry, K. T. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature*, *463*(7283), 899.

9. Krijgsman, O., Carvalho, B., Meijer, G. A., Steenbergen, R. D., & Ylstra, B. (2014). Focal chromosomal copy number aberrations in cancer—Needles in a genome haystack. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1843*(11), 2698-2704.

10. Taylor, A. M., Shih, J., Ha, G., Gao, G. F., Zhang, X., Berger, A. C., ... & Lazar, A. J. (2018). Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell*, *33*(4), 676-689.

11. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545-15550.

12. Tesson, B. M., Breitling, R., & Jansen, R. C. (2010). DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC bioinformatics*, *11*(1), 497.

13. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., & Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*, *160*(1-2), 48-61.

14. Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., ... & Matthews, L. (2015). The reactome pathway knowledgebase. *Nucleic acids research*, *44*(D1), D481-D487.

15. Bezu, L., Sauvat, A., Humeau, J., Gomes-da-Silva, L. C., Iribarren, K., Forveille, S., ... & Senovilla, L. (2018). eIF2α phosphorylation is pathognomonic for immunogenic cell death. *Cell Death & Differentiation*, 1.

16. Jauhiainen, A., Thomsen, C., Strömbom, L., Grundevik, P., Andersson, C., Danielsson, A., ... & Åman, P. (2012). Distinct cytoplasmic and nuclear functions of the stress induced protein DDIT3/CHOP/GADD153. *PloS one*, *7*(4), e33208.