Robust Assessment of Genomic Imbalance in Diffuse Large B-Cell Lymphoma Confirms Inferior Outcome Is Associated with Genomic Complexity and Identifies Potential Therapeutic Pathway Targets

Lizalynn Dias[1], PhD, Venkata Thodima[1],PhD, Julia Friedman[1], PhD, Asha Guttapalli[1], Geetu Mendiratta[1], PhD, Sergei Syrbu, MD[2], and Jane Houldsworth[1], PhD

[1] Cancer Genetics, Inc., Rutherford, NJ, [2]University of Iowa, Iowa City, IA

Introduction

Diffuse large B-cell lymphomas (DLBCL) display marked clinical, pathologic, and genetic heterogeneity. With current frontline immunotherapy (RCHOP), only about 40% of patients are cured, with most relapses occurring within the first 2-3 years. Patients are currently risk-stratified based primarily on clinical features where the inclusion of molecular biomarkers into risk assessment could impact the potential to identify those patients most likely to have refractory disease or have an early relapse. Various cytogenomic studies have revealed the prognostic significance of genomic gain/loss in DLBCL, but their lack of utility and reproducibility across datasets can be attributed to not only different patient populations, but also the use of disparate platforms and analytical methods. The goal of the present study was to use a common analytical approach across different clinical datasets, to identify common regions of copy number aberrations (CNAs) in DLBCL and genomic loci with robust prognostic value in DLBCL.

Datasets

The present study involved the analysis of raw SNP or aCGH data from the following publically available *in-silico* (IS) and in-house (IH) datasets:

IS-172 (GSE11318)
- i. 172 fresh frozen DLBCL tumor biopsy specimens newly diagnosed CHOP treated patients (Ref. 1).
- ii. Whole Genome 385K Tiling arrays (NimbleGen)
- iii. Gene expression data was available for 162 of the 172 specimens profiled in the IS-172 dataset

IS-169 (GSE34171)
- i. 169 frozen biopsy specimens from newly diagnosed RCHOP (N=70) and CHOP treated patients (Ref. 2)
- ii. SNP array 6.0 (Affymetrix)

IS-51HR (E-MEXP-3463)
- i. 51 fresh frozen DLBCL tumor specimens from high risk (HR) i.e. high-IPI RCHOP treated patients
- ii. Human Genome CGH Microarray (4x180K) array (Agilent)

IS-124 (GSE15127)
- i. 124 frozen tumor specimens taken at diagnosis from RCHOP treated patients (Ref.3)
- ii. Mapping 250K Nsp SNP Array (Affymetrix)

IH-41
- i. 41 FFPE diagnostic blocks from patients treated with doxorubicin- and rituximab-containing regimens
- ii. Custom designed targeted 4x44K oligonucleotide aCGH (Agilent)(Ref. 4)

Overall Strategy

***In silico* publically available datasets**

↓ GISTIC

32 regions commonly altered in at least 2 of the 3 datasets (16 gains, 17 deletions)

↓

36 minimal common regions (MCRs) mapped within the 32 regions (19 gains, 17 deletions)

↓

50 aberrations (Abr.) mapped within the 36 MCRs (27 gains, 23 deletions, when present GISTIC peak included in CNA)

↓

Establishment of robust scoring criteria for application across platforms

↓ Validation and application of scoring criteria

- **Correlate CNA with expression profiles and pathway analysis (IS-172)**
- **Correlate CNA with overall survival in RCHOP and RCHOP-like treated patients (IS-169(N =70), IS-124, IH-41)**
- **Correlate genomic complexity with overall survival in RCHOP and RCHOP-like treated patients (IS-169(N =70), IS-124)**

Results

In-silico Identification of Overlapping CNAs in DLBCL

GISTIC on datasets: IS-172, IS-169 and IS-51HR

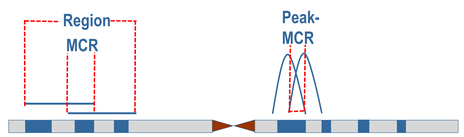


Figure 1- After applying the GISTIC algorithm to three available *in silico* datasets, overlapping regions and minimal common regions (MCR) were identified within either larger chromosomal regions or peak regions in at least 2 of the 3 datasets.

Aberrations (CNAs) and Calling Criteria

MCR	Abr.	Location	Chr.	MCR Start	MCR Stop	Peak(Y/N)
1	1-3	A1q22.1-q25.1	1	145,390,170	174,990,852	Y
2	4	A1q31.3-q32.1	1	196,256,151	206,325,234	N
3	5	A2p16.1-p15	2	57,798,324	64,087,027	Y
4	6	A3q27.3-q29	3	187,651,865	196,853,350	Y
5	7	A5p15.33	5	304,462	1,944,706	N
6	8	A6p21.32-p21.2	6	32,192,560	37,798,373	Y
7	9	A7p22.2	7	2,804,495	3,921,089	N
8	10	A8q24.3	8	144,439,451	146,138,827	N
9	11	A9q34.3	9	138,543,735	140,878,804	N
10	12	A11p15.4	11	3,225,556	3,808,818	N
11	13	A11q23.3	11	120,599,996	120,820,140	Y
12	14	A12q13.11-q13.12	12	46,766,301	51,194,543	N
13	15	A12q14.2-q21.1	12	64,762,188	73,252,451	Y
14	16	A13q31.3	13	91,878,316	92,275,215	Y
15	17	A16q24.3	16	89,644,837	89,662,874	N
16	18-24	A18p11.21-18q23	18	15,381,304	77,861,995	Y
17	25	A19p13.3	19	260,999	2,183,174	N
18	26	A19q13.33-q13.43	19	50,132,339	58,866,674	N
19	27	A21q22.3	21	43,014,315	47,970,581	Y

Table 1A- 19 MCRs of Gain

MCR	Abr.	Location	Chr.	MCR Start	MCR Stop	Peak(Y/N)
20	28	D1p36.32-36.31	1	2,436,018	6,342,694	Y
21	29	D1p13.1	1	116,368,309	117,144,417	Y
22	30	D2q22.3	2	144,608,496	146,483,034	N
23	31	D2q24.2	2	161,836,066	161,963,707	N
24	32	D3p21.31-p21.2	3	48,849,803	50,713,840	N
25	33	D3p14.2	3	60,408,342	60,690,266	N
26	34	D6p21.33	6	30,919,878	31,536,224	Y
27	35-39	D6q11.1-q27	6	61,962,715	170,373,079	Y
28	40	D8p22-21.3	8	13,241,832	23,112,533	Y
29	41	D9p24.1	9	4,971,499	5,596,997	Y
30	42	D9p21.3	9	21,448,157	22,622,538	Y
31	43	D10q23.31	10	90,567,045	90,986,391	Y
32	44	D13q14.13-q14.3	13	46,917,541	53,172,272	N
33	45-46	D15q15.1-q21.1	15	40,295,857	46,224,648	Y
34	47	D16q12.1-q12.2	16	50,674,936	52,984,751	N
35	48-49	D17p13.3-p11.2	17	1,000,000	16,936,602	Y
36	50	D19p13.3	19	6,427,364	6,851,788	Y

Table 1B- 17 MCRs of Loss

Validation of Calling Criteria with Cell of Origin (IS-172)

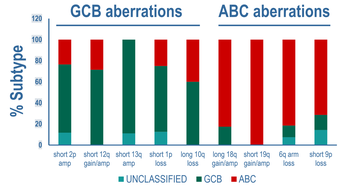


Figure 2- Alterations previously reported to be enriched in cell-of-origin DLBCL subtypes (Ref. 1) were similarly tested for association after applying the newly established scoring criteria.

Results

Integrated Analysis of CNA and Expression

Raw expression data for 162 specimens of IS-172 were RMA-normalized and regional genes were correlated with scored CNAs by univariate t-test. Genes exhibiting at least a 1.2 fold change and $P \leq 0.05$ after FDR were considered further.

The expression of a total of 569 unique genes correlated with at least one aberration. Of these, significant gene expression correlation was:

- Not found for 17 abrs
- Was found for 16 abrs with peaks, but none mapped to the peak
- Was found for 11 abrs with peaks, and mapped to the peak (see below)
- Was found for 6 abrs with no peak, and mapped in MCR (see below)

MCR	MCR (#genes)	Peak (#genes)	Gene(s)
A1q22.1-q25.1	67	1	LOC100505773
A1q31.3-q32.1	7	N/A	RABIF, TMEM183, ZC3H11A, SNRPE, MDM4, ELK4, NUCKS1
A2p16.1-p15	12	1	PAPOLG
A3q27.3-q29	27	6	SDHAP2, PCYT1A, TFRC, LOC440993, MUC20, SDHAP2
A7p22.2	1	N/A	GNA12
A8q24.3	28	N/A	NFKBIL2, CPSF1, FOXH1, PARP10, …
A12q13.11-q13.12	8	N/A	TMEM106C, SENP1, PFKM, ASB8, ZNF641, TROAP, MCRS1, RACGAP1
A12q14.2-q21.1	9	1	CNOT2
A18p11.21-18q23	75	1	MEX3C
A18p11.21-18q23	81	2	FECH, NARS
A19q13.33-q13.43	46	N/A	BCL2L12, SPIB, PRMT1, PPP2R1A, CNOT3, …
D6q11.1-q27	119	1	TMEM30A
D6q11.1-q27	153	1	TNFAIP3
D13q14.13-q14.3	12	N/A	NEK3, SETDB2, VPS36, DHRS12, …
D15q15.1-q21.1	9	1	UBR1
D15q15.1-q21.1	9	1	TRIM69
D17p13.3-p11.2	51	3	POLR2A, SENP3, FXR2

Table 2- Genes whose expression correlated with CNAs.

Pathway Analysis

Ingenuity pathway analysis was performed on the total 569 unique genes which correlated with a CNA. Five canonical pathways were significantly enriched (FDR $p \leq 0.05$).

Pathway	P-value	Genes Involved
p53 Signaling	p = 0.01	TP53,PMAIP1,PIK3CA,TOPBP1,MDM2, PIK3R4,GML,BCL2,MDM4,STAG1,PIK3C3,ATR,TNFRSF10A,CDK2
PKCθ Signaling in T Lymphocytes	p = 0.025	MAP2K4,FYN,PIK3CA,MAP3K13,PPP3CC,MALT1,MAP3K5,MAP3K4,PIK3R4,NFATC1,POU2F1,MAP3K7,PIK3C3,VAV1
Geranylgeranyldiphosphate Biosynthesis	p = 0.029	FDPS,COX10,GGPS1
B Cell Receptor Signaling	p = 0.039	MAP2K4,FCGR2C,PIK3CA,MAP3K13,MAP3K4,PPP3CC,MALT1,MAP3K5,PIK3R4,FCGR2B,NFATC1,SYNJ2,MAP3K7,PIK3C3,VAV1,INPP5K
RANK Signaling in Osteoclasts	p = 0.039	MAP2K4,PIK3CA,MAP3K7,PIK3C3,MAP3K13,TAB2,PPP3CC,MAP3K5,MAP3K4,PIK3R4,NFATC1

Table 3- Pathways significantly enriched after IPA.

Results

Association of CNAs with Patient Outcome

Aberration	IS-169 (N=70)	IS-124 (N=124)	IH-41 (N=41)
Gain			
12q13.11-q13.12	p ≤ 0.05	n.s.	n.s.
12q14.2-q21.1	p ≤ 0.05	n.s.	n.s.
16q244	p ≤ 0.01	n.s.	n.s.
19q13	p ≤ 0.01	n.s.	n.s.
Loss			
2q24.2	p ≤ 0.01	n.s.	n.s.
6q21q11.1	n.s.	n.s.	p ≤ 0.01
8p22-21.3	n.s.	p ≤ 0.05	n.s.
9p21.3	n.s.	n.s.	p ≤ 0.01
15q15.1-q21.2	n.s.	p ≤ 0.05	n.s.
17p13.3-11.2	p ≤ 0.05	n.s.	n.s.

Table 4- Using the datasets of patients treated with either RCHOP- or RCHOP-like regimens, individual aberrations were tested by the log rank statistic for association with overall survival. Of the 50 identified aberrations, 10 were found to be significantly associated with adverse outcome, but not across all datasets.

Correlation of Clinically Relevant Aberrations

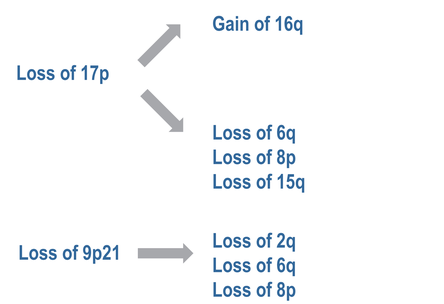
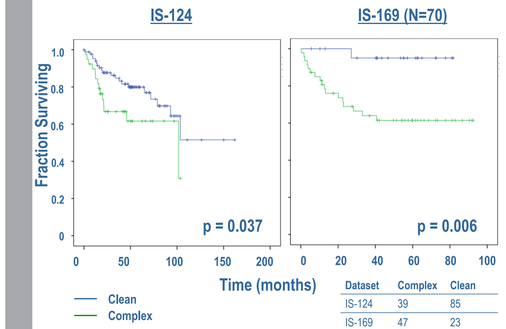


Figure 3- Aberration-Aberration correlation. Of the 10 aberrations found to be significantly correlated with disease outcome (Table 4), pairwise correlation was performed across all datasets to detect the presence or absence of co-occurring genomic alterations.

Measurement of Complexity and Association with Outcome

Method 1- Complexity based on total number of detectable MCRs within RCHOP and RCHOP-like treated specimens.

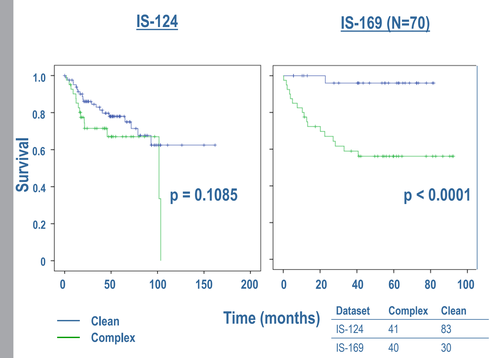


Dataset	Complex	Clean
IS-124	39	85
IS-169	47	23

Figure 4- Kaplan-Meier plot of dataset IS-124 and IS-169(N=70) which were all RCHOP treated patients. The median number of detectable aberrations across all *in silico* datasets was determined to be 1, therefore, any specimen with ≥2 aberrations was called as complex, and any with ≤1 were called as clean.

Results

Method 2- Complexity based on having an aberration within the CDKN2A-TP53-RB-E2F axis (Ref. 2)



Dataset	Complex	Clean
IS-124	41	83
IS-169	40	30

Figure 5- Kaplan-Meier plot of dataset IS-124 and IS-169(N=70) which were all RCHOP treated patients. Samples were scored as complex if they contained any one of aberrations 2 or 3, 8, 9, 15, 26, 42, 44, 47, and 48 or 49. These aberrations were previously shown to harbor genes important in the CDKN2A-TP53-RB1-E2F axis. If these aberrations were not detected, samples were called clean. **statistical significance was found in the original dataset IS-169*

Conclusions

- The application of GISTIC to genomic profiling datasets generated from various platforms allowed the delineation of regions of common genomic imbalance in DLBCL.
- Of the 32 overlapping regions of genomic gain/loss observed in at least 2 of the 3 datasets, 36 MCRs were found. 50 individual aberrations were further delineated based on peaks identified by GISTIC.
- Robust scoring criteria were established for each of the 50 aberrations, that were platform agnostic.
- Correlation of genomic imbalance with gene expression profiling data revealed several genes previously shown to have a role in DLBCL, such as TNFAIP3 and SPIB.
- Pathway analysis of genes whose expression correlated with CNAs verified previously known affected pathways in DLBCL, such as p53 signaling and B Cell-Receptor signaling and identified other pathways that could represent novel therapeutic targets.
- Of the 50 aberrations, 10 were found to be associated with shorter overall survival, though not reproducible across datasets.
- Association of genomic complexity with patient outcome was dependent on the method by which complexity was assessed.

References

1. Lenz G., Wright G.W., Emre, N.C., et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. PNAS, 2008. .
2. Monti S, Chapuy B, Takeyama K, et al. . *Integrative analysis reveals an outcome-associated and targetable pattern of p53 and cell cycle deregulation in diffuse large B cell lymphoma*. Cancer cell 2012;22:359-372
3. Scandurra M, Mian M, Greiner TC, et al. . Genomic lesions associated with a different clinical outcome in diffuse large B-Cell lymphoma treated with R-CHOP-21. British journal of haematology 2010
4. Houldsworth J, Guttapalli A, Thodima V, et al. . *Genomic imbalance defines three prognostic groups for risk stratification of patients with chronic lymphocytic leukemia*. Leukemia & lymphoma 2013.

Conflict of Interest

L.D., V.T., J.F., A.G., G.M., and J.H. are stock/stock options holders. S.S. is not affiliated.