Cross-Platform Assessment of Genomic Imbalance in Diffuse Large B-Cell Lymphoma Identifies Candidate Novel Loci and Genes with Prognostic Value and Roles in Lymphomagenesis.

Lizalynn Dias[1], Venkata Thodima[1], Geetu Mendiratta[1], Asha Guttapalli[1], Camille Gonzalez[2], Jocelyn Maragulia[3], Andrew D, Zelenetz[3], Julie Teruya-Feldstein[2], R.S.K. Chaganti[3,4], Jane Houldsworth[1]

[1] Cancer Genetics, Inc., Rutherford, NJ, [2] Department of Pathology, [3] Department of Medicine, and [4] Cell Biology Program, Memorial Sloan-Kettering Cancer Center

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
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-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-180 (GSE34171)

i. 180 frozen biopsy specimens from newly diagnosed RCHOP and CHOP treated patients
ii. SNP array 6.0 (Affymetrix)

IS-124 (GSE15127)

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

IH-46

i. 46 CD20 enriched FFPE core punch specimens from RCHOP treated patients with de-novo DLBCL
ii. Custom designed targeted 4x44K oligonucleotide array (Agilent)

Methods

DNA isolation and array-CGH

Genomic DNA was isolated from 1-3 FFPE core-punches (1.5mm diameter x 2.3mm length) of the IH-46 dataset following the protocol by EH van Beers et al. [1]. Labeled DNA was hybridized to a custom designed targeted 4x44K oligonucleotide array consisting of genomic regions designed to assist in the prognosis of various mature B-cell neoplasms including DLBCL [2].

GISTIC Analysis and Aberration Detection

Recurrent CNAs were identified by applying Genomic Identification of Significant Targets In Cancer (GISTIC) to CBS segmented data from the in-silico datasets IS-172 and IS-51HR (≥ 0.2 for gain and ≤ -0.2 for loss, FDR q-value of 0.25). Raw aCGH/SNP data were loaded into the Nexus Copy Number Analysis Software 6.1 where aberration detection was performed using the Rank segmentation algorithm.

Clinical Correlative Analyses

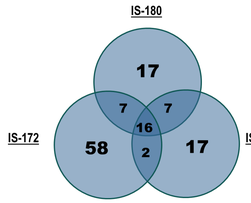
The clinical correlation of CNAs and genomic complexity to was assessed in RCHOP datasets IH-46, IS-124 and IS-51HR using log-rank statistic (*P*-value ≤ 0.05). Wherever applicable the Beniamini Hochberg multiple testing correction was performed.

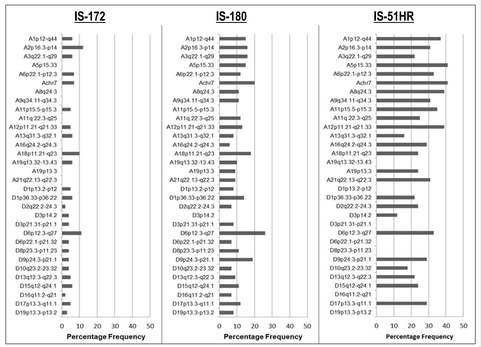
Results

In-silico Identification of Overlapping CNAs in DLBCL

GISTIC identified regions of significant CNAs and peak-CNAs in the DLBCL datasets IS-180 [3],IS-172 and IS-51HR.

- A CNA region of overlap was defined as the largest region of genomic gain/loss including all overlapping regions across least 2 of the 3 datasets.
- The smallest region of overlap within a CNA region was defined as its region-MCR.
- A peak-CNA of overlap was defined as the largest region of genomic gain/loss including all overlapping GISTIC peaks across least 2 of the 3 datasets.
- The smallest region of overlap within the peak-CNA was defied as its peak-MCR.





Average frequency=6% Average frequency=12% Average frequency=26%

The frequency of the identified 32 regions was found to be very different among datasets and not surprisingly, the dataset comprising the high IPI patients (IS-51HR) exhibited a higher percentage frequencies for all aberrations.

CNAs Associated with Clinical Outcome

35 MCRs were tested for association with overall survival in R-CHOP treated cohorts

Aberration	OS *P*-value/Corrected *P*-value		
	IH-46 (N=46)	IS-124 (N=124)	IS-51HR (N=49)
Gain			
3q27.3-q29	<0.001/<0.001	n.s.	n.s.
9q34.3	n.s.	n.s.	<0.001/0.011
19p13.3	n.s.	n.s.	0.038/n.s.
Loss			
1p13.1	0.046/n.s.	n.s.	n.s.
8p22-21.3	n.s.	0.024/n.s.	n.s.
9p24.1	0.007/n.s.	n.s.	n.s.
13q14.13-q14.3	0.029/n.s.	n.s.	n.s.
15q12-q24.1	n.s.	n.s.	<0.001/0.012
17p13.3-p11.2	0.007/0.050	n.s.	n.s.

Novel loci associated with poor outcome, n.s.- not significant

Delineation of Peak-MCRs and Identification of Gene

Genes present in overlapping peak-MCRs

Region			Peak-MCRs	
Cyto Band	Start	End	#	Genes
A2p16.3-p14	60,993,696	61,147,244	2	*REL**, *PAPOLG*
A3q22.1-q29	195,262,873	196,054,003	15	*TRFC (CD71)**, PCYT1A, *MUC20, MUC4, MIR570*
A6p22.1-p12.3	32,690,831	32,709,466	1	***HLA-DQA2***
Achr7	1,017,099	1,081,180	3	*CYP2W1, C7orf50, MIR339*
	101,059,950	101,481,721	3	*COL26A1, MYL10, CUX1*
A9q34.11-q34.3	138,543,735	140,343,466	81	*CAD9, NOTCH1, EGLF7, MIR4673, MIR4674*
A12p11.21-q21.33	70,345,927	70,719,344	1	*CNOT2*
A13q31.3-q32.1	92,020,763	92,032,734	0	No genes
A21q22.13-q22.3	43,959,706	44,358,680	5	*SLC37A1, PDF9A, WDR4, NDVF3, HERV-Fb1*
D1p36.33-p36.22	2,926,556	3,762,888	12	*TP73, PRDM16, MIR551A*
D1p13.2-p12	117,070,416	117,111,959	1	***CD58*****
D6p22.1-p21.32	31,167,499	31,326,959	3	*HCG27, HLA-B, HLA-C*
D6p12.3-q27	138,049,196	138,241,766	1	***TNFAIP3*****
D8p23.3-p11.23	19,362,769	19,667,874	1	***CSGALNACT1***
D9p24.3-p21.1	5,106,681	5,109,576	1	***JAK2***
	21,978,347	21,984,369	2	*CDKN2A**, *MTAP*
D10q23.2-23.32	90,597,629	90,775,628	5	*FAS**, *FAS-AS1, ANKRD22*, *STAMBPL1, ACTA2*
D15q12-q24.1	43,133,828	43,370,177	2	*TTBK2, UBR1*
	44,939,261	45,026,122	4	*SPG11, PATL2, B2M, TRIM69*
D17p13.3-q11.1	7,841,724	7,920,934	2	*CNTROB, GUCY2D*
D19p13.3-p13.2	6535980	6620909	1	***CD70*****

Genes present in non-overlapping peak-CNAs

Region			Peak-CNAs	
Cyto Band	Start	End	#	Individual peaks
A3q22.1-q29	187,458,723	187,526,170	1	***BCL6***
A5p15.33	1,269,366	1,286,296	1	***TERT***
A11p15.5-p15.3	9,617,734	9,759,805	1	***SWAP70***
A16q24.2-q24.3	89,644,837	89,662,874	1	***CPNE7***
A18p11.21-q23	60,862,909	60877603	1	*BCL2**
	53,061,763	53,421,274	2	*TCF4**
A19p13.3	2,60,999	2,183,174	80	*MUM1**
D3p14.2	60,436,146	60,495,719	1	***FHIT***
D6p12.3-q27	106,425,102	106,579,472	1	***PRDM1*****
D10q23.2-23.32	89,533,586	89,729,441	4	*PTEN*
D13q12.3-q22.3	50,600,024	51,380,221	6	DLEU1, DLEU2, MIR16-1

*Gene implicated in DLBCL
Correlate with expression
Sole genes at peaks

Overlapping peak-MCRs mapped to single as well as multiple gene loci.

The identification of genes with known roles in DLBCL like *REL*, *FAS*, *TNFAIP3*, *CD70*, *CDKN2A* at overlapping peak-MCRs of both deletion and amplification confirm the validity of the analysis approach used and very likely identifies novel gene targets with previously unacknowledged roles in the pathogenesis of DLBCL.

Specimen diversity between datasets was highlighted by the identification of peak-CNAs in individual datasets without any overlap with the others. These peak-CNAs mapped to known oncogenes and genes implicated in lyphomagenesis.

Assessment of Genomic Complexity

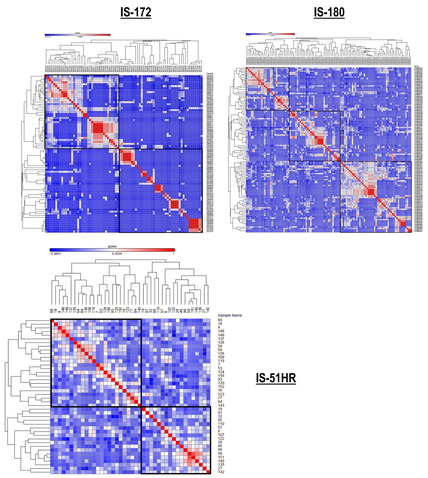
Genomic complexity was assessed by 3 different methods and tested for outcome association in three R-CHOP treated datasets (IH-46, IS-124, IS-51HR).

1. ***CDKN2A-TP53-RB-E2F* axis**
The presence of at least one of CNAs identified to effect the *CDKN2A-TP53-RB-E2F* axis namely; gain of 1q23.3, 6p21.32, 7q22.1, 12q15, 19q13.42, or loss of 9p21.3, 13q14.2, 17p13.1 and 16q12.2 [3].

	CDKN2A-TP53-RB-E2F		
Dataset	% Complex	% Non-complex	*P*-value
IS-51HR (n=49)	67	33	n.s.
IS-124 (n=124)	23	77	n.s.
IH-46 (n=46)	59	41	n.s.

n.s.- not significant

2. **Associated CNAs**
Hierarchical clustering of 35 MCR CNAs across datasets (IS-172, IS-51HR and IS-180) identified co-occurring aberrations that "marked" complexity namely; gains of 3q, 11q, 18, 19q and loss of 1p, 2q, 3p, 6q, 9p, 17p,19p.



	Associated CNAs		
Dataset	% Complex	% Non-Complex	*P*-Value
IS-51HR (n=49)	59	41	n.s
IS-124 (n=124)	37	63	n.s.
IH-46 (n=46)	54	46	n.s

n.s.- not significant

3. **Absolute number of CNAs**
Taking into consideration only the identified recurring 35 region-MCRs, the median was calculated for the total number of aberrations for all specimens in the IS-124, IS-172, IS-180, and IH-46 datasets. Median=1. Samples with more than 1 of the 35 MCRs were considered complex.

	Absolute number of CNAs		
Dataset	% Complex	% Non-Complex	*P*-Value
IS-51HR (n=49)	80	20	n.s
IS-124 (n=124)	31	69	0.0732
IH-46 (n=46)	83	17	n.s

n.s.- not significant

CNA complexity assessed by the three methods mentioned above failed to correlate with outcome in the RCHOP datasets tested.

Integrated Analysis of CNA and Expression

For the 162 samples of the IS-172 dataset with available gene expression data genes that positively correlated with copy number status of the 35 region-MCRs were identified using the "affy" Bioconductor (R) package. Gene ontology and pathway analysis was performed within Ingenuity.

Sixteen of the 35 region-MCRs were found to contain a total of 396 differentially expressed genes showing positive correlation to copy number status (P≤0.05, FDR).

Gene ontology (GO): (Molecular and Cellular Functions)
Cell death and survival (56)
DNA replication, recombination, repair (33)
Cell cycle (27)

Pathway analysis

Canonical Pathway	*P*-value	*P*-value (FDR)	#	Total	Genes
B Cell Receptor Signaling	<0.001	0.046	11	162	*SYNJ2, MAP2K4, MAPK14, MAP3K7, PIK3C3, PPP3CC, MALT1, INPP5K, MAP3K4, NFATC1, PTEN*
RANK Signaling in Osteoclasts	<0.001	0.046	8	92	*MAP2K4, MAPK14, MAP3K7, PIK3C3, TAB2, PPP3CC, MAP3K4, NFATC1*
p53 Signaling (apoptosis)	0.001	0.051	8	95	*TP53, PMAIP1, MAPK14, PIK3C3, MDM2, TNFRSF10A, PTEN, BCL2*
Regulation of IL-2 Expression in Activated and Anergic T Lymphocytes	0.001	0.068	7	84	*MAP2K4, SMAD2, FYN, SMAD4, PPP3CC, MALT1, NFATC1*
TGF-β Signaling	0.001	0.078	7	89	*MAP2K4, SMAD2, MAPK14, MAP3K7, FOXH1, SMAD4, BCL2*
PKCθ Signaling in T Lymphocytes	0.002	0.087	8	127	*MAP2K4, FYN, MAP3K7, PIK3C3, PPP3CC, MALT1, MAP3K4, NFATC1*
RAR Activation	0.002	0.087	10	175	*MAP2K4, SMAD2, MAPK14, PNRC1, GTF2H5, SMAD4, NCOR1, RDH13, CITED2, PTEN*
Mouse Embryonic Stem Cell Pluripotency	0.003	0.095	7	99	*TP53, TCF4, MAPK14, MAP3K7, PIK3C3, SMAD4, DVL2*
PEDF Signaling	0.003	0.095	6	77	*TP53, TCF4, MAPK14, PIK3C3, BCL2*

Conclusions

Using a platform-agnostic approach, common and novel loci of genomic imbalance in DLBCL were identified, of which some were found to have potential clinical significance and could be included, with additional validation, in patient risk assessment.

This analysis also afforded the delineation of peak-MCRs and the identification of novel genes with roles in lymphomagenesis representing potential therapeutic targets in DLBCL.

Genomic complexity assessed in three different ways failed to significantly correlate with overall survival in the datasets tested.

References

1. van Beers EH, Joosse SA, Ligtenberg MJ, et al. . *A multiplex PCR predictor for aCGH success of FFPE samples*. Br J Cancer 2006;94:333-337.

2. 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.

3. 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.