Location of the t(4;14) translocation breakpoint within the NSD2 gene identifies a subset of high-risk NDMM patients

Stong, N et al.

Supplemental Methods and Data File

Methods

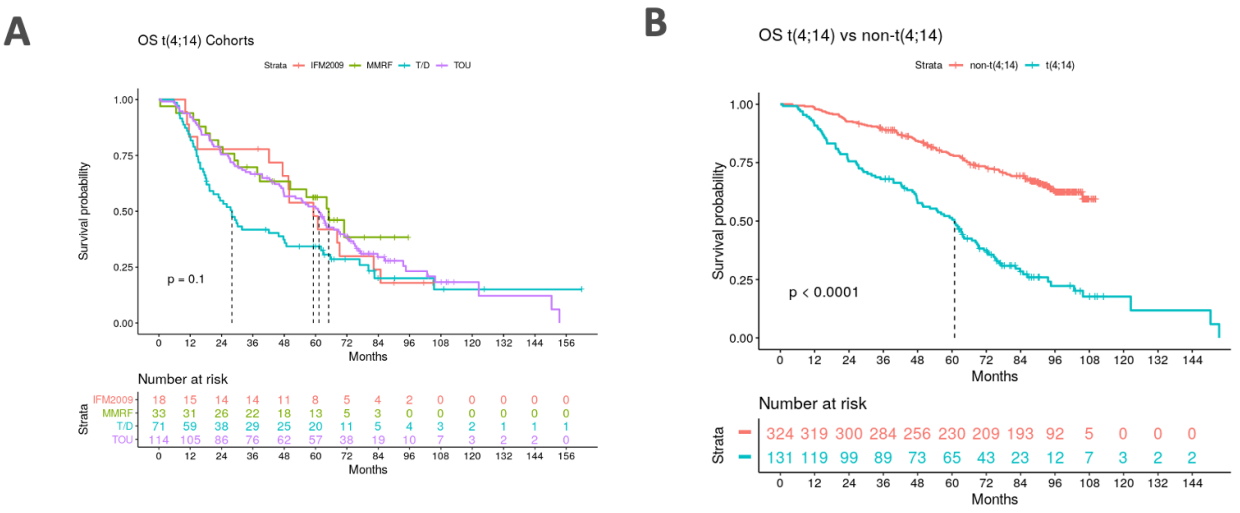
Computational Methods

FASTQ files were aligned to hg38 using BWA-mem. Base recalibration of alignments was performed using GATK v4. Single nucleotide variants (SNVs) and small insertions/deletions (indels) were called using MuTect2 (v4) and annotated with ANNOVAR. Matched normal samples were available. To confirm that tumor and normal samples were taken from the same individual, kinship coefficient testing was performed.[1] Cytogenetic assessments from WGS were made by MANTA and used to identify translocation DNA breakpoint location. Copy number aberration (CNA) calling was performed using Battenberg as previously described.[2-4] RNASeq reads were mapped and gene expression was quantitated using STAR and Salmon.[5,6] Fusion transcripts were analyzed using STAR-fusion for RNASeq.[7,8]

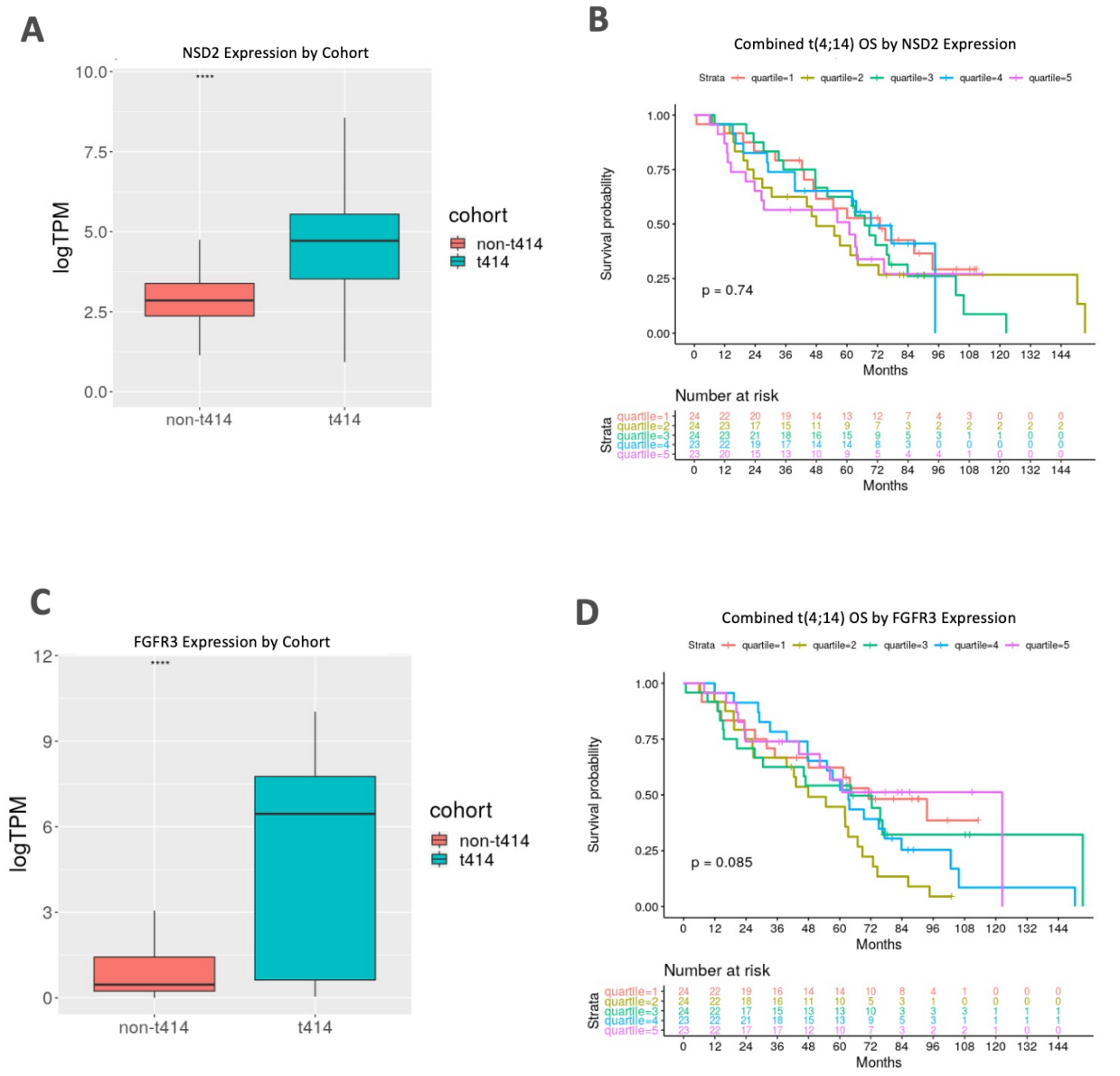
Statistical Methods

Kaplan Meier (KM) plots were generated using survival package in R and p-values were calculated using Cox proportional hazards regression. Fisher tests (for CNA/mutations comparison) and t-test for gene expression comparison were performed.

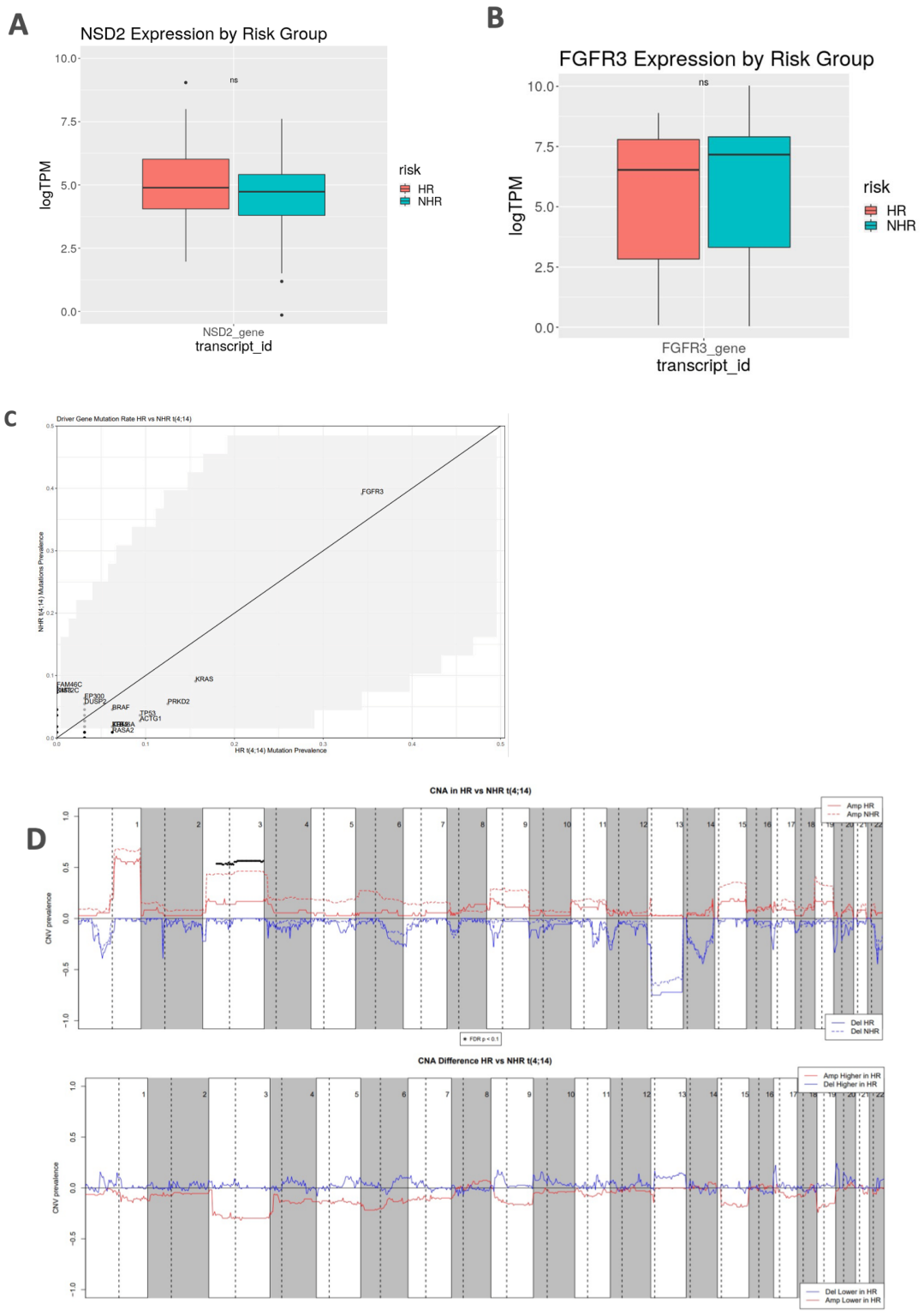
Supplemental Tables and Figures



Supplementary Figure 1. **A)** Overall survival Kaplan-Meier (KM) plot of all t(4;14) cohorts. **B)** Survival KM plot of the discovery t(4;14) cohort compared to non-t(4;14) patients cohort from IFM-2009

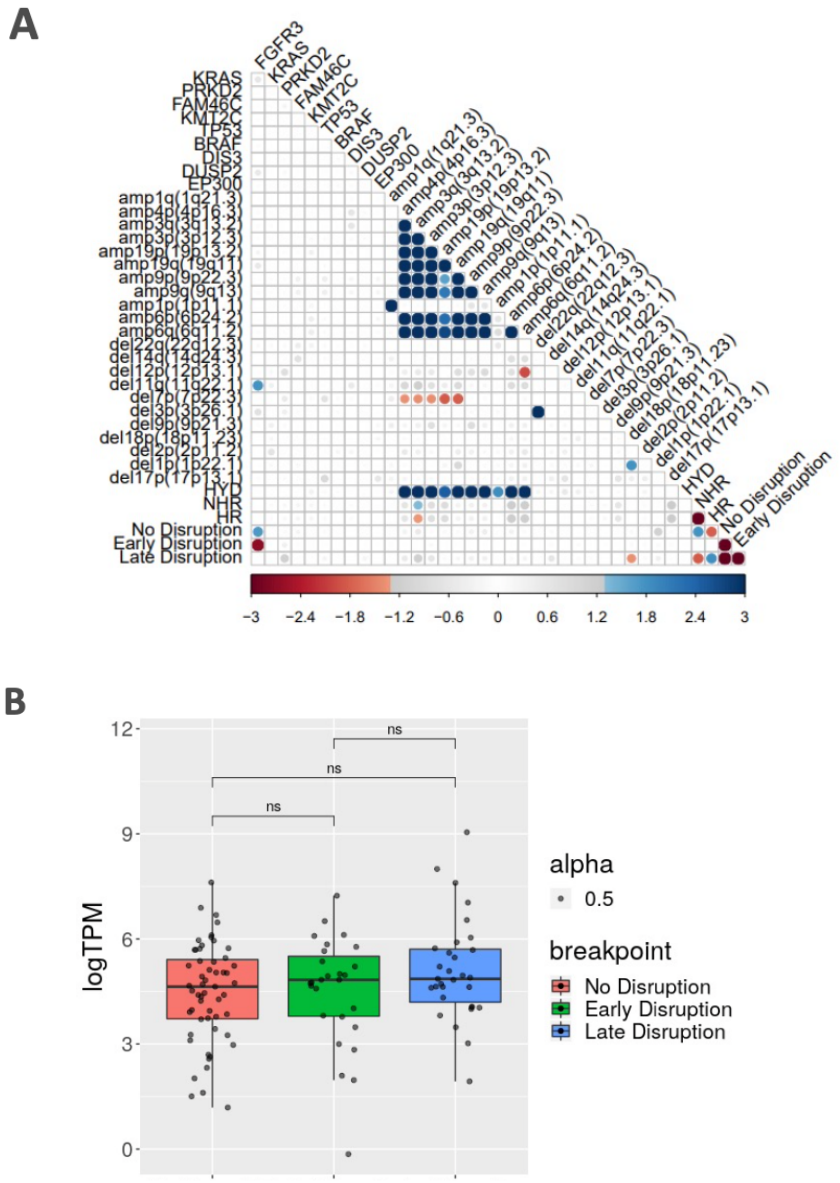


Supplementary Figure 2: **A)** NSD2 expression in t(4;14) versus non-t(4;14). **B)** overall survival based on NSD2 expression level. **C)** FGFR3 expression in t(4;14) versus non-t(4;14). **D)** overall survival based on FGFR3 expression level.

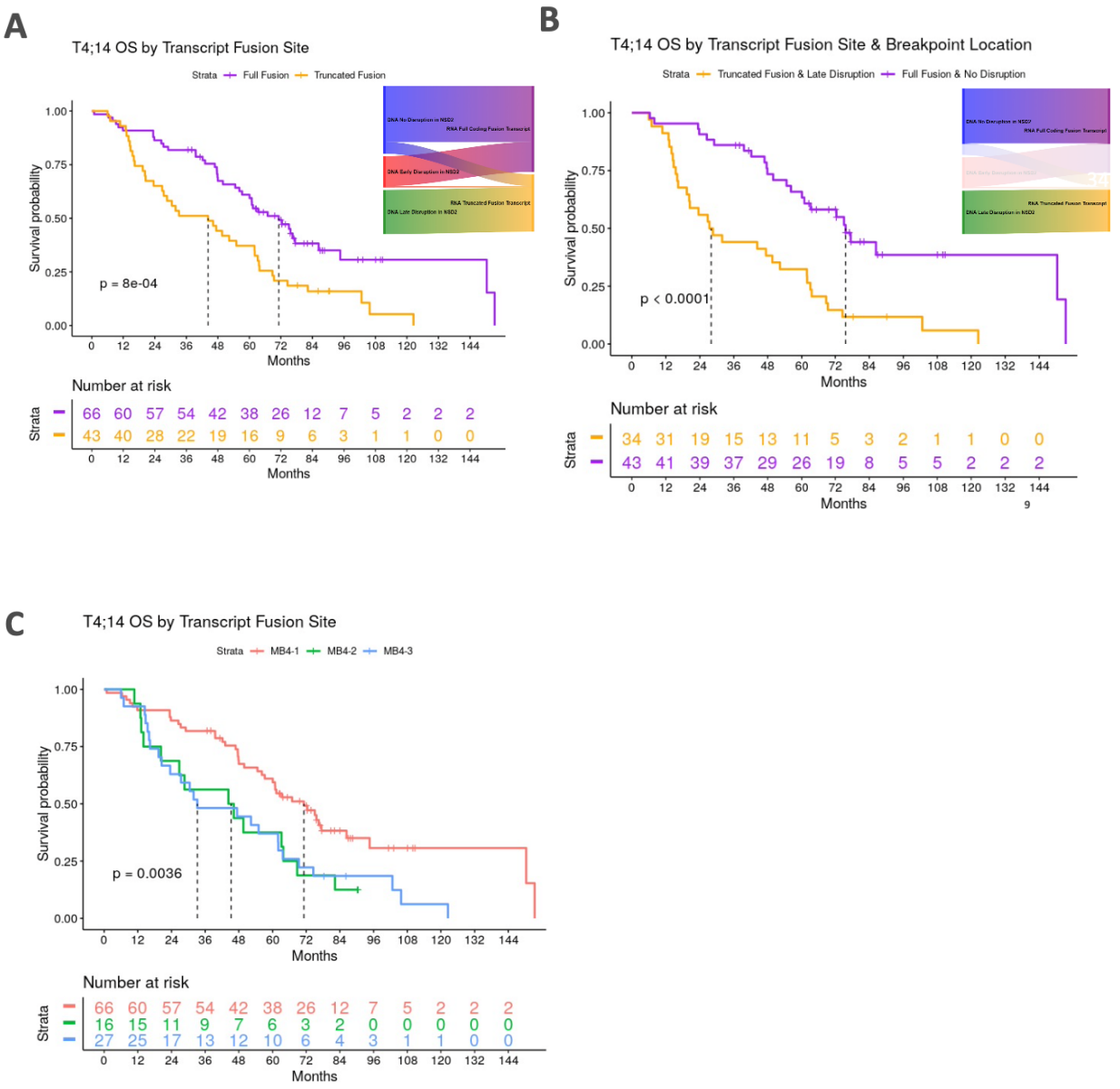


Supplementary Figure 3: A) NSD2 expression in high-risk vs non-high-risk t(4;14) patients. **B)** As in A, but for FGFR3. **C)** Driver mutations in HR t(4;14) vs NHR patients. Gray shaded zone indicates non significance between the two groups. **D)** Top: Copy number aberrations between HR and NHR t(4;14) patients.

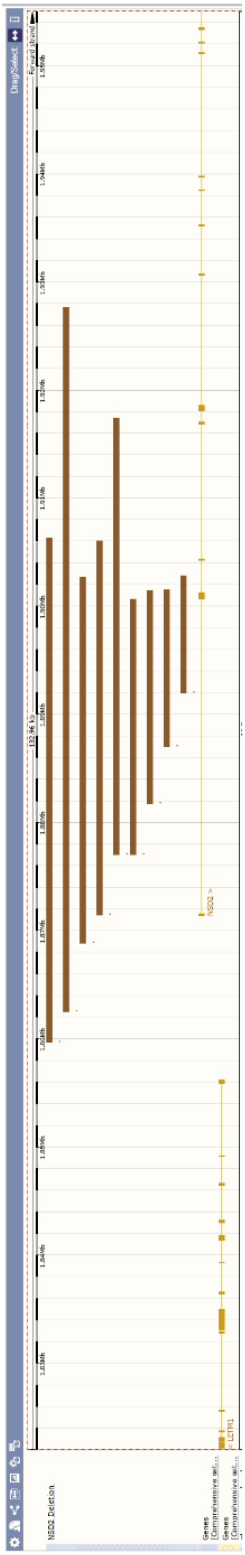
Increased copy number shown in red, decreases in blue. Solid lines = HR, dashed lines = NHR. Bottom: difference in CNA between HR and NHR groups.



Supplementary Figure 4: A) Genomic and risk feature co-occurrence within t(4;14) including breakpoint location. **B)** NSD2 expression by risk group based on breakpoint location.



Supplementary Figure 5: Categorization of patients based on expressed NSD2 isoform and associated overall survival. **A)** Overall survival of patients expressing a full-length NSD2 transcript vs truncated fusion transcript. **B)** Overall survival of patients with a full-length NSD2 transcript and *no-disruption* breakpoint (blue region of inset Sankey plot) vs patients with a truncated NSD2 transcript and a *late-disruption* (green/yellow region of inset Sankey plot). Patients with an *early-disruption* are not shown. **C)** overall survival of t(4;14) patients by transcript fusion site (MB4-1 vs MB4-2 vs MB4-3).



Supplementary Figure 6: Patients with a *no-disruption* breakpoint location who express a short NSD2 transcript due to presence of a deletion. Deleted region is shown for individual patients.

1a. Gain. Most significant band per arm

Band	fdr_p	Significant Genes of Interest in Arm/Band
1q44	1.69E-09	
3p26.2	1.68E-15	CRBN(3p26.1)
3q12.1	1.22E-02	
4p15.33	7.83E-09	FGFR3(4p16.3)
5p15.1	2.64E-13	
5q31.2	2.97E-14	
7p15.2	1.35E-07	
7q11.21	1.81E-07	BRAF(7q34), KMT2C(7q36.1)
8q24.21	1.74E-03	
9p24.2	1.69E-20	
9q34.3	4.07E-12	
11p14.2	2.73E-11	
11q13.3	1.60E-16	BIRC2(11q22.1)
13p11.1	6.31E-07	
15q11.2	3.53E-09	
18p11.22	1.80E-03	
18q12.1	5.33E-03	
19p13.11	1.51E-10	
19q11	8.03E-10	PRKD2(19q13.32)

1b. Deletion. Most significant band per arm

band	fdr_p	Significant Genes of Interest in Arm/Band
1p22.1	2.73E-02	
2q31.1	3.66E-02	
3p25.3	8.50E-06	CRBN(3p26.1)
4p15.32	2.18E-03	
4q24	4.01E-02	
5p15.32	4.07E-02	
6p25.3	4.01E-02	
7p21.3	9.31E-05	
8p23.2	7.72E-04	
9p21.3	8.08E-05	
11q22.1	2.53E-06	BIRC2(11q22.1)
12p13.1	5.19E-06	KRAS(12p12.1)
13p11.1	7.76E-10	
13q12.12	1.62E-09	DIS3(13q21.33)
14q31.1	1.24E-05	
16q24.3	2.45E-05	
18p11.31	1.33E-04	
18q11.1	4.01E-02	
20p11.22	7.93E-03	
20q11.1	1.18E-02	

Significant differences are shown in both directions (eg, significant for the t(4;14) pts and also significant differences for the non-t(4;14) pts. Refer to Figure 1D for visualization.

Supplementary Table 1: t(4;14) CNA prevalence differences between t(4;14) and non-t(4;14). Significant differences are shown in both directions (eg, significant for the t(4;14) pts and also significant differences for the non-t(4;14) pts. Refer to Figure 1D for visualization.

Feature	HR Prevalence	NHR Prevalence	P value (fisher exact)
FGFR3 Mutation	34.4%	39.1%	0.68
1qAMP	22.2%	22.7%	1

Supplementary Table 2: Prevalence of FGFR3 mutation and 1q amplification (≥4 copies) in HR and NHR groups.

Univariate analysis, OS					
Feature	Coefficient	Exp (coefficient)	SE (coefficient)	z	Pr(>|z|)
Early-disruption breakpoint	0.4532	1.573339	0.269162	1.683746	0.09
Late-disruption breakpoint	0.946426	2.576485	0.23577	4.014187	6.00E-05
Age	0.007393	1.007421	0.011883	0.622172	0.53
ISS2	0.208831	1.232237	0.325723	0.641131	0.52
ISS3	0.474668	1.60748	0.330303	1.437068	0.15
del17p	0.639053	1.894686	0.287903	2.219681	0.03
del1p	0.991787	2.696049	0.428768	2.313108	0.02
Amp1q	0.366131	1.442144	0.253422	1.444746	0.15
Gain 1q	0.142486	1.153137	0.211962	0.672224	0.5
Trisomy chr 5	-0.86605	0.420608	0.395539	-2.18955	0.03
Trisomy chr 21	NA	NA	0	NA	NA
Multivariate analysis, OS					
Feature	Coefficient	Exp (coefficient)	SE (coefficient)	z	Pr(>|z|)
Early-disruption breakpoint	0.620361	1.8596	0.308976	2.007799	0.04
Late-disruption breakpoint	0.786256	2.195163	0.296841	2.648746	8.10E-03
Age	0.013795	1.01389	0.013232	1.04256	0.3
ISS2	0.404269	1.498207	0.358886	1.126455	0.26
ISS3	0.109541	1.115765	0.383638	0.285531	0.78
del17p	1.132696	3.104012	0.363609	3.115145	1.80E-03
del1p	1.383721	3.989722	0.470666	2.939921	3.30E-03
Amp1q	0.256835	1.292832	0.309147	0.830785	0.41
Gain 1q	0.279212	1.322088	0.295521	0.944813	0.34
Trisomy chr 5	-0.77854	0.459076	0.485201	-1.60457	0.11
Trisomy chr 21	NA	NA	0	NA	NA

Supplementary Table 3: OS Association with Breakpoint Location and Other High-Risk Features

Patient distribution by breakpoint group and landmark analysis with an OS cutoff of 24 months				
	No Disruption	Early Disruption	Late Disruption	Total
HR	15.6% (5)	28.1% (9)	56.2% (18)	100.0% (32)
NHR	54.5% (54)	22.2% (22)	23.2% (23)	100.0% (99)
Patient distribution by breakpoint group and landmark analysis with an OS cutoff of 36 months				
	No Disruption	Early Disruption	Late Disruption	Total
HR	21.4% (9)	23.8% (10)	54.8% (23)	100.0% (42)
NHR	56.7% (51)	23.3% (21)	20.0% (18)	100.0% (90)

Supplementary Table 4: Prevalence of patients in breakpoint groups split by differing overall survival cutoff points for definition of the high-risk group.

References for Supplementary Content

1. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. *Bioinformatics (Oxford, England)*. 2010;26(22):2867-2873.
2. Nik-Zainal S, Alexandrov Ludmil B, Wedge David C, et al. Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell*. 2012;149(5):979-993.
3. Gooding S, Ansari-Pour N, Towfic F, et al. Multiple Cereblon genetic changes associate with acquired resistance to Lenalidomide or Pomalidomide in Multiple Myeloma. *Blood*. 2020.
4. Cun Y, Yang T-P, Achter V, Lang U, Peifer M. Copy-number analysis and inference of subclonal populations in cancer genomes using Sclust. *Nature Protocols*. 2018;13(6):1488-1501.
5. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2012;29(1):15-21.
6. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods*. 2017;14(4):417-419.
7. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology*. 2013;31(3):213-219.
8. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*. 2010;38(16):e164-e164.