Supplemental Information

ZFN and TALEN sequences

ZFN:

The recognition helices for ZFN^{EWS} are as follows:

ZFN	F1	F2	F3	F4	F5	F6
ZFN^{EWS-R}	QSNHRKT	DRSDLSR	RSDNLTR	RSDYLST	QRSHRNT	
ZFN^{EWS-L}	DRSHLSR	QSGHLSR	RSDHLSQ	TSANRTT	QSGALAR	TSSNRKT
ZFN^{FLI-R}	RSDNLSV	QKATRIN	DQSNLRA	QSGHLQR	QSGHLSR	
ZFN^{FLI-L}	RSDTLSE	DRSNLTR	LKQNLDA	QSGSLTR	RSDNLTT	

Each ZFN was designed to have a distinct obligate heterodimeric architecture through modification of the FokI nuclease domain (Doyon et al. 2011). These modifications decrease the potential off-target sites and promote correct heterodimer formation while reducing formation of incorrect heterodimers and homodimers.

We used obligate heterodimeric domains of FokI (Doyon et al. 2011): ELD (ZFN^{EWS-R}) /KKR (ZFN^{EWS-L}) and RDD (ZFN^{FLI-R})/DRR (ZFN^{FLI-L}).

TALEN:

TALENs were assembled by a method derived from Huang et al. (Huang et al. 2011). For each TALEN subunit, the fragment containing the 16 RVD segment was obtained from single unit vectors: A (NI), T (NG), G (NN) and C (HD), kindly provided by the laboratory of Bo Zhang. The assembled TALENs were subcloned in the pCS2 vector for *ALK* and pVax vector for *NPM1*. Sequences of encoded TALEN proteins are as follow (underline: TAL recognition domain; *italic*: wt FokI domain). Where indicated, obligate heterodimeric domains of FokI were used: ELD (TAL^{NPM-L}) /KKR (TAL^{NPM-L}) and RDD (TAL^{ALK-R})/DRR (TAL^{ALK-L}).

$TALEN^{ALK-L}$:

MAPKKKRKVYPYDVPDYAGYPYDVPDYAGSYPYDVPDYAAHGTVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGH
GFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKI
AKRGGVTAVEAVHAWRNALTGAPLNLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQA
LETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLP
VLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLT
PAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS
NGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALE
TVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVL

CQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPE
QVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIAS**NI**GGRPALESIVAQLSRPDPALAALTNDHLVALACL
GGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVA*GSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDR*
ILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHI
NPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGE
INFRS

TAL$^{ALK-R.}$:

MAPKKKRKVYPYDVPDYAGYPYDVPDYAGSYPYDVPDYAAHGTVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGH
GFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKI
AKRGGVTAVEAVHAWRNALTGAPLNLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQA
LETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLP
VLCQAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLT
PAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS
NGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALE
TVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVL
CQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPE
QVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGRPALESIVAQLSRPDPALAALTNDHLVALACL
GGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVA*GSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDR*
ILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHI
NPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGE
INFRS

TALEN^{NPM-L}:

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTH
AHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPGTLDTGQLLKIAKRG
GVTAVEAVHAWRNALTGAPLNLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETV
QRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQ
AHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQV
VAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGG
KQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR
LLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQAH
GLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVA
IASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRP
ALDAVKKGLPHAPALIKRTNRRIPERTSHRVA*GSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEM*
KVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNE
WWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINFR
S

TAL^{NPM-R}:

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPMVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTH

AHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPGTLDTGQLLKIAKRG
GVTAVEAVHAWRNALTGAPLNLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETV
QRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQ
AHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQV
VAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGG
KQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQR
LLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAH
GLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVA
IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRP
ALDAVKKGLPHAPALIKRTNRRIPERTSHRVA*GSQLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEM*
KVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNE
WWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

Supplemental Figures Legends

Supplemental Figure S1

A. Map of *EWSR1-FLI1* patient breakpoint junctions in *EWSR1* intron 7 and in a 1.5 kb region from FLI1 intron 5 from (Zucman-Rossi et al. 1998). *EWSR1* intron 7 is 1.5 kb; *FLI1* intron 5 is 23.3 kb. Locations of the ZFN binding sites are indicated by the boxes and the scissors. Translocations for which breakpoint junctions for both derivative chromosomes have been sequenced but are not adjacent to each other are specified for both der(22) and der(11).
B. Sequences of the segments shown in A. The tumor breakpoints for each junction are indicated by a slash in blue. The ZFN binding sites are highlighted in yellow. Breakpoints falling within the ZFN sites are in red.

Supplemental Figure S2

Der(22) and der(11) translocation junction sequences obtained in hES-MP cells after combined ZFN^{EWS} and ZFN^{FLI} expression. Junctions were obtained from individual wells after ZFN expression from 3 independent experiments; as such, all junctions likely arose from independent events. Junctions are arranged according to the total deletion from the DSBs. The chromosome 22 sequence at the breakpoint is in black and the chromosome 11 sequence is in red, both shown with fill-in of the 5' overhang and with the ZFN binding sequence underlined. The cleavage sites are likely to be variable as to whether they include the terminal nucleotides shown in parenthesis, i.e., T and A for der(22) or T and G for der(11). The number of deleted bases from each end (dots) is indicated. Microhomologies are underlined. Junctions seen repetitively that occur in the ZFN overhangs at 1 bp microhomology are indicated by the asterisks. Inserted sequences are indicated in green. If the deletion extends beyond the chromosome sequence indicated at the top of the figure, a few bp flanking the deletion are indicated, with microhomology underlined. Junctions with long insertions are indicated at the end, along with the inserted sequence and its derivation(s).

Supplemental Figure S3

Breakpoint junction characteristics from patients with Ewing sarcoma tumors (EWTUM) reported in Zucman-Rossi et al. (Zucman-Rossi et al. 1998) (GenBank accession numbers AJ229253 to AJ229365). Deletions and duplications (insertions) from chromosomes 11 and 22 were previously analyzed. Microhomology at the junctions of both der(11) and der(22) as well as additional insertions not from chromosomes 11 and 22 were analyzed for the current report.

Supplemental Figure S4

Sequences of complex t(11;22)(q24;q12) breakpoint junctions with insertions obtained with ZFN expression in hES-MP cells. Junctions are derived from different sets of experiments and include some junctions from Supplemental Fig. S1 that are also described in Fig. 2. Duplications from chromosome 11 and 22 are shown in red and black; insertions from other sources are green. Microhomologies between the derivative chromosomes and the inserted sequences are underlined. Total length and DNA origins of insertions are also reported.

Supplemental Figure S5

TALEN-induced translocations in RPE-1 cells.

A. TALEN cleavage activity in RPE-1 cells, as monitored by the T7-endonuclease assay. Controls include untransfected cells (wt) and cells expressing only the left part of TAL^{ALK} ($TAL^{ALK\text{-}L}$).

B. Nested PCR to detect derivative chromosomes der(2) and der(5) in RPE-1 cells. Translocation breakpoint junctions are only detected after co-expression of TAL^{NPM} and TAL^{ALK}.

C. Single round PCR to detect derivative chromosomes der(5) in RPE-1 cells. Translocation breakpoint junctions are detected after expression of both TAL^{NPM} and TAL^{ALK} by PCR of the fragment marked (*) in Fig. 3C on serial dilutions of genomic DNA (5, 2.5, 1.25, 0.625, and 0.312 ng). The number of times the PCR was positive for each dilution from 3 total experiments is indicated. The 951 bp marker corresponds to a der(5) junctions without end modification. The smaller fragment seen in the first lane likely correspond to a junction with a large deletion.

D. RT-PCR detection of the *NPM1-ALK* fusion transcript after TAL^{NPM} and TAL^{ALK} co-expression in RPE-1 cells.

E. Western blot detection of the NPM1-ALK fusion protein in RPE-1 cells after TAL^{NPM} and TAL^{ALK} co-expression. The antibody is directed against ALK.

Supplemental Figure S6

In vitro mapping of TALEN cleavage sites. DNA oligonucleotides of 68 bp containing TAL^{NPM} (**A**) and TAL^{ALK} (**B**) binding sites were ^{32}P-labelled at the 5' end on either the top (left) or bottom (right) strand and then annealed to the complementary unlabelled strand. Cleavage reactions were carried out with in vitro translated TALENs. The positions of the spacer sequences for the labeled strands, as deduced from the G+A sequencing, are shown to the side of the gel, as are

the positions of the left and right binding sites of the TALENs (blue and red bars, respectively). The major cleavage sites are represented by arrows on the sequences below the gels.

Supplemental Figure S7

Translocation junction sequences for t(2;5)(p23;q35) obtained in Jurkat (**A**) and hES-MP (**B**) cells after combined TAL^{NPM} and TAL^{ALK} expression. Junctions were obtained from individual wells after TALEN expression from 3 independent experiments; as such, all junctions likely arose from independent events. Junctions are arranged according to the total deletion from the DSBs. The chromosome 5 sequence at the breakpoint is in black and the chromosome 2 sequence is in red, both shown with fill-in of the 5' overhang and with the TALEN binding sequence underlined. Cleavage sites are likely to be variable (Supplemental Fig. S6): Nucleotides indicated in italics represent the maximal 5' overhang considering both top strand and bottom strand cleavage, while the terminal nucleotides in parentheses represent the maximal 5' overhang considering just bottom strand cleavage. Deleted bases from each end are indicated by dots. Microhomologies are underlined and inserted sequences not clearly derived from chromosomes 5 or 2 are indicated in green. Junctions with long insertions are at the end, along with the inserted sequence and its derivation(s).

Supplemental Figure S8

Induction of t(2;5)(p23;q35) translocations in Jurkat cells with heterodimeric TALENs.

A. ALCL translocations induced by TALENs with heterodimeric FokI domains.

Top: Nested PCR was performed to detect derivative chromosomes der(2) and der(5) in Jurkat cells. Translocation breakpoint junctions are only detected after expression of both TAL^{NPM} and TAL^{ALK}.

Bottom: Single round PCR detects after heterodimeric TALEN expression detects der(5) in Jurkat cells, as in Fig. 3E, but at an apparent slightly lower frequency than homodimeric TALENs. In serial dilutions of genomic DNA, the der(5) breakpoint junction was obtained in 2 of 3 experiments with 12.5 ng DNA (representing ~2000 cells), implying that translocations arose at a frequency of $\sim 0.5 \times 10^{-3}$.

B. Translocation junction sequences obtained after expression of TAL^{NPM} and TAL^{ALK} containing heterdimeric FokI domains.

Supplemental Figure S9

Chr(5) and Chr(2) reversed breakpoint sequences obtained from ALCL patient-derived cell lines SUDHL-1 and SUPM2. Junction sequences were obtained after cloning and are presented as in Supplemental Figure S7.