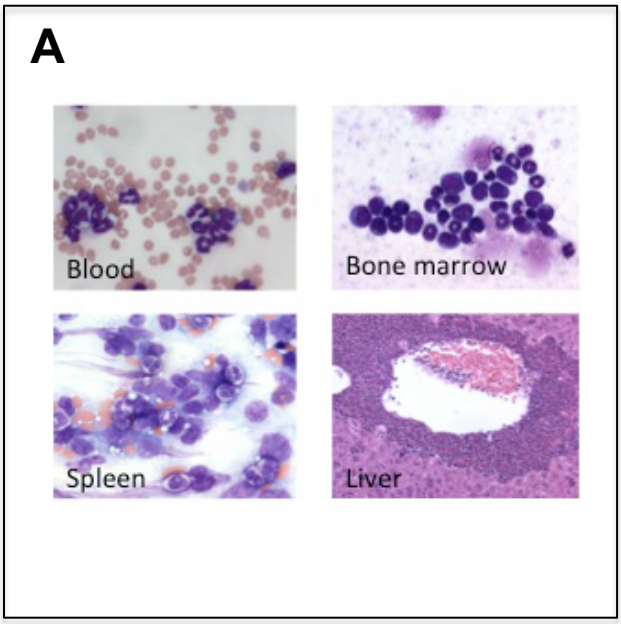
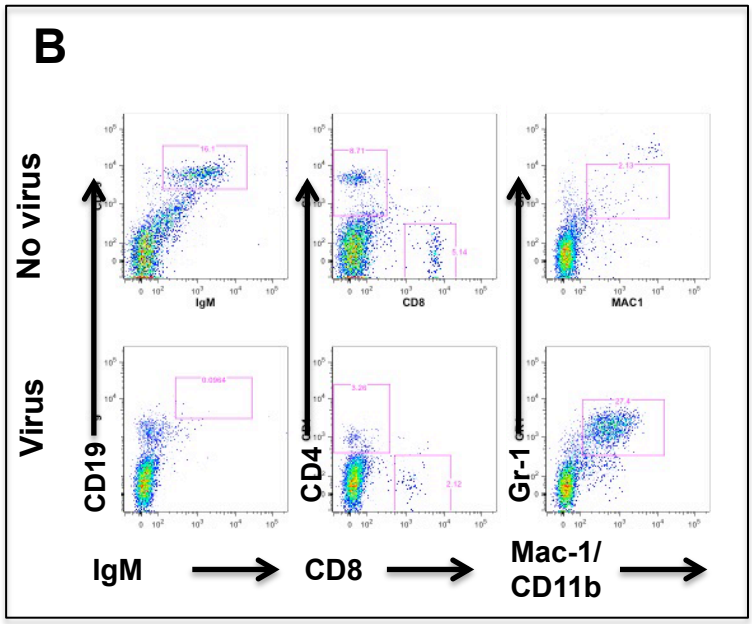
Retroviral insertional mutagenesis identifies the del(5q) genes, *CXXC5*, *TIFAB* and *ETF1*, as well as the Wnt pathway, as potential targets in del(5q) myeloid neoplasms

Angela Stoddart,[1] Zhijian Qian,[2] Anthony A. Fernald,[1] Rachel J. Bergerson,[3] Jianghong Wang,[1] Theodore Karrison,[4,5] John Anastasi,[5,6] Elizabeth T. Bartom,[7] Aaron L. Sarver,[3] Megan E. McNerney,[5,8] David A. Largaespada,[3] and Michelle M. Le Beau[1,5]

[1]Department of Medicine, University of Chicago, Chicago, IL; [2]Department of Medicine, and University of Illinois Cancer Center, University of Illinois at Chicago, Chicago, IL; [3]Department of Pediatrics, Masonic Cancer Center, University of Minnesota, Minneapolis, MN; [4]Department of Public Health Sciences, University of Chicago, Chicago, IL; [5]University of Chicago Medicine Comprehensive Cancer Center, Chicago, IL: [6]Department of Pathology, University of Chicago, Chicago, IL; [7]Department of Biochemistry and Molecular Genetics, Northwestern University Feinburg School of Medicine, Chicago, IL; and [8]Departments of Pathology and Pediatrics, and Institute for Genomics and Systems Biology, University of Chicago, Chicago, IL, USA

Correspondence: astoddar@bsd.uchicago.edu
doi:10.3324/haematol.2015.139527





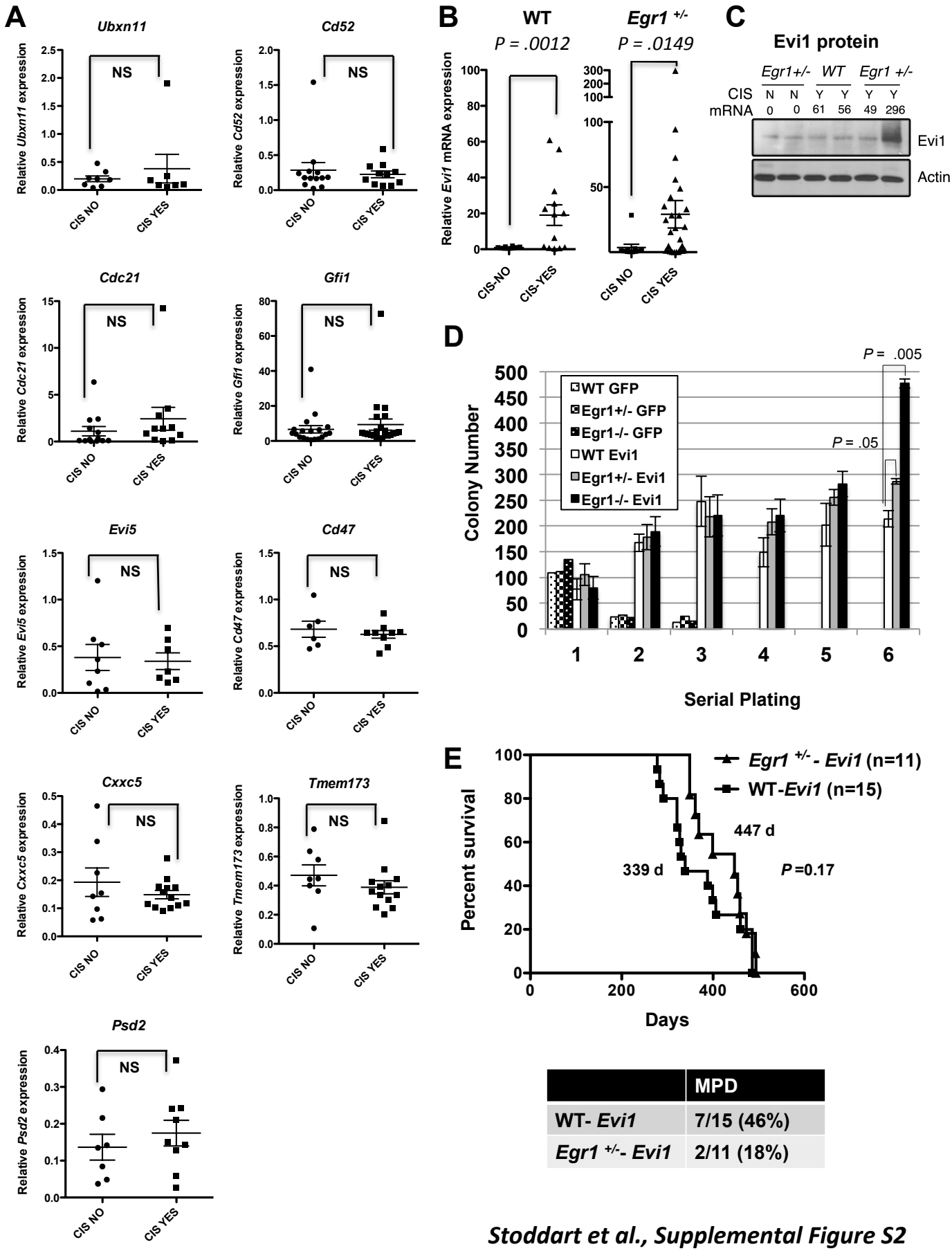
C

Donor/Recipient	WBC (K/uL)	RBC (M/uL)	% Gr1+ Mac1+ in spleen	% Gr1- KIT+ in spleen	Spleen (g)	Survival (days)
Donor: 2938	>200	4.9	14	12	0.9	424
Recip: 3811	>200	3.8	8	6	0.8	33
Recip: 3812	146*	7.9*	nd	nd	0.6	32
Recip: 3813	>200	2.7	9	12	0.9	33
Donor: 2489	13.8	6.3	4	66	1.1	375
Recip: 3682	nd	nd	0.1	48	0.6	74
Recip: 3683	173.4	1.8	0.4	69	0.7	119
Recip: 3684	nd	nd	nd	nd	0.6	73
Donor: 2495	139.2	6.6	29	9	0.8	263
Recip: 3585	133.6	6.2	17	6	0.6	35
Recip:3586	91.4	6.2	17	7	0.4	36
Recip:3814	41.2	3.9	17	7	0.5	40
Recip: 3815	31.9	2.9	33	29	0.4	40

* CBC taken 6 days prior to sacrifice

Stoddart et al., Supplemental Figure S1

Supplemental Figure S1. MOL4070LTR-treated WT and *Egr1*$^{+/-}$ mice develop myeloid neoplasms. (A) Peripheral blood and bone marrow smears, spleen touch preparation stained with Wright-Giemsa, and liver section stained with hematoxylin and eosin from a typical *Egr1*$^{+/-}$ mouse that developed a myeloid neoplasm with maturation. Infiltration of mature myeloid cells into liver, a non-hematopoietic tissue. **(B)** Flow cytometric analysis of B cells ($CD19^+$ IgM^+), T cells ($CD4^+$ or $CD8^+$), and myeloid cells ($Gr1^+Mac1^+$) in spleens from an untreated *Egr1*$^{+/-}$ mouse and a representative *Egr1*$^{+/-}$ mouse with a myeloid neoplasm. **(C)** Three million spleen cells from primary diseased mice were transplanted into sublethally irradiated recipient mice. Survival of secondary recipients was significantly reduced, and the phenotype recapitulates that of the donor mouse.



	MPD
WT- *Evi1*	7/15 (46%)
Egr1 $^{+/-}$ *- Evi1*	2/11 (18%)

Stoddart et al., Supplemental Figure S2

Figure 2. Evaluation of cooperation of overexpression of *Evi1* with *Egr1* haploinsufficiency. (A) Expression of candidate genes was measured in spleen cells isolated from diseased *Egr1*$^{+/-}$ mice by real-time RT-PCR and normalized to *Gapdh.* A comparison of gene expression in spleen cells isolated from diseased mice with a proviral insertion proximal to candidate gene (CIS YES) versus expression in cells isolated from diseased mice without the corresponding proviral insertion (CIS NO) reveals that proviral integration did not significantly increase or decrease gene expression (P> 0.05 and not signficant (NS) for 9 candidate genes)**. (B)** mRNA expression of *Evi1* was measured in the spleen of diseased WT and *Egr1*$^{+/-}$ mice by real-time RT-PCR, and normalized to *Gapdh*. Elevated *Evi1* expression is seen in myeloid neoplasms from WT and *Egr1*$^{+/-}$ mice with proviral integrations proximal to *Evi1 (CIS YES)* compared to mice without a CIS proximal to *Evi1* (CIS NO). **(C)** Protein expression of Evi1 was measured by Western blot analysis of spleen cells isolated from WT and *Egr1*$^{+/-}$ mice that either had a proviral CIS near *Evi1* (Y) or not (N). The relative increase in *Evi1* mRNA expression is shown for comparison. Evi1 protein was only observed at high levels in mouse 2489, which showed an ~300 fold increase in Evi1 by Q-PCR (shown in B). Actin protein is shown as a loading control. **(D)** *Egr1*$^{+/-}$ and *Egr1*$^{-/-}$ bone marrow cells with enforced *Evi1* expression (or control empty vector expressing only GFP) display enhanced replating capacity compared to WT controls, when serially plated in methocult with IL-3, IL-6, SCF and Epo. After 12 days of incubation, colonies of greater than 50 cells were scored for morphology by inverted light microscopy, cells were harvested, serially replated and scored. **(E)** WT and *Egr1*$^{+/-}$ progenitors transduced with *Evi1,* and transplanted into lethally irradiated WT recipients show similar survival indicating that *Evi1* does not cooperate with *Egr1* haploinsufficiency *in vivo*. In fact, a larger proportion of WT mice than *Egr1*$^{+/-}$ mice developed an MPD. Most of the Egr1$^{+/-}$ died at >1 year post transplantation due to non-hematological effects (see results).

Supplemental Table S1. GREAT statistical analysis of the MSigDB pathway terms that significantly associate with putative target genes in WT and *Egr1*$^{+/-}$ mice.

Term Name-enriched in WT[1]	Rank	Raw P-Value	FDR Q-Value	Fold Enrichment
Cancer				
Chronic myeloid leukemia	4	9.28288e-7	2.04223e-4	4.2639
Acute myeloid leukemia	6	3.08511e-6	4.52482e-4	4.6618
Glioma	13	2.11304e-5	1.43036e-3	4.1846
Endometrial cancer	15	2.85012e-5	1.67207e-3	4.0634
Non-small cell lung cancer	18	3.57350e-5	1.74704e-3	4.2721
Cadmium induces DNA synthesis and proliferation in macrophages	24	1.06335e-4	3.89895e-3	6.6934
MAPK signaling				
MAPK signaling pathway (KEGG)	1	1.39500e-8	1.22760e-5	2.8058
MAPKinase Signaling Pathway	2	1.10085e-7	4.84375e-5	4.4580
JAK-STAT signaling				
p38 MAPK Signaling Pathway	14	2.60153e-5	1.63525e-3	4.4166
Jak-STAT signaling pathway	7	6.40471e-6	8.05164e-4	3.1313
Cytokine signaling				
Chemokine signaling pathway	17	3.13268e-5	1.62162e-3	2.8971
IL-2 Receptor Beta Chain in T cell Activation	3	2.83160e-7	8.30603e-5	6.9264
Genes involved in TRAF6 Mediated Induction of the antiviral cytokine IFN-a cascade	12	1.98369e-5	1.45470e-3	4.2105
Immune system signaling				
Members of the BCR signaling pathway	19	4.77179e-5	2.21009e-3	4.4970
Genes involved in Signaling in Immune system	11	1.12699e-5	9.01595e-4	2.3096
Tp53 Signaling				
Tp53 signaling pathway	8	7.66814e-6	8.43496e-4	4.0353
Innate Immunity				
Genes involved in Toll Receptor Cascades	5	2.57460e-6	4.53129e-4	3.5960
Genes involved in MAP kinases activation in TLR cascade	9	9.40307e-6	9.19411e-4	4.5259
Genes involved in Toll Like Receptor 3 (TLR3) Cascade	16	2.90950e-5	1.60023e-3	4.0551
Genes involved in Innate Immunity Signaling	10	9.70231e-6	8.53803e-4	3.2700

[1] The test set of 520 genomic regions picked 764 (4%) of all 20,221 genes. *MSigDB Pathway* has 880 terms covering 5,986 (30%) of all 20,221 genes, and 37,583 term - gene associations. 880 ontology terms (100%) were tested using an annotation count range of [1, Inf].

Term Name-enriched in *Egr1*$^{+/-}$ [2]	Rank	Raw P-Value	FDR Q-Value	Fold Enrichment
Cancer				
Chronic myeloid leukemia	1	5.84751e-13	5.14581e-10	5.0539
Acute myeloid leukemia	24	9.76591e-4	3.58084e-2	2.8138
MAPK signaling				
MAPK signaling pathway (KEGG)	3	5.47237e-8	1.60523e-5	2.3382
MAPKinase Signaling Pathway	27	1.20013e-3	3.91152e-2	2.4402
JAK-STAT signaling				
Jak-STAT signaling pathway	4	3.72387e-7	8.19250e-5	2.9077
Cytokine signaling				
Cytokine-cytokine receptor interaction	12	1.04490e-4	7.66262e-3	2.0671
Genes related to IL4 receptor signaling in B lymphocytes	31	1.38848e-3	3.94149e-2	3.4850
Immune system signaling				
B Cell Antigen Receptor	5	4.34568e-6	7.64839e-4	5.3187
Members of the BCR signaling pathway	19	5.23407e-4	2.42420e-2	3.1887
Genes involved in Signaling in Immune system	6	1.68836e-5	2.47626e-3	2.0017
PIP3 signaling				
Genes related to PIP3 signaling in B lymphocytes	9	5.14616e-5	5.03180e-3	3.8427
Genes related to PIP3 signaling in cardiac myocytes	22	8.53305e-4	3.41322e-2	2.5233
Wnt signaling				
Genes related to Wnt-mediated signal transduction	29	1.30216e-3	3.95137e-2	2.1690
Hematopoiesis				
Hematopoietic cell lineage	2	6.84355e-10	3.01116e-7	4.7865
Monocyte and its Surface Molecules	10	7.26492e-5	6.39313e-3	8.8933
Wound Healing				
Genes involved in Hemostasis	7	2.22020e-5	2.79110e-3	2.0366
Genes involved in Cell surface interactions at the vascular wall	15	3.44862e-4	2.02319e-2	2.5659

[2] The test set of 800 genomic regions picked 1,169 (6%) of all 20,221 genes. *MSigDB Pathway* has 880 terms covering 5,986 (30%) of all 20,221 genes, and 37,583 term - gene associations. 880 ontology terms (100%) were tested using an annotation count range of [1, Inf].

Supplemental Table S2. GREAT statistical analysis of the GO Molecular Function[1] terms that significantly associate with putative target genes in WT and *Egr1*$^{+/-}$ mice.

Genotype	Term Name[2]	Rank	Raw P-Value	FDR Q-Val	Fold Enrichment
WT	Transcription Regulatory Region DNA Binding	6	2.48577e-9	1.34811e-6	2.5558
WT	Regulatory Region DNA Binding	7	3.99681e-9	1.85795e-6	2.5182
WT	Purinergic Receptor Activity	22	1.78663e-5	2.64259e-3	7.3545
WT	Adenosine Receptor Activity, G-Protein Coupled	49	4.85844e-4	3.22640e-2	20.2037
Egr1 $^{+/-}$	Kinase Binding	14	2.11070e-6	4.90587e-4	2.0418
Egr1 $^{+/-}$	Protein Kinase Binding	16	2.85179e-6	5.79983e-4	2.0914
Egr1 $^{+/-}$	Cytokine Binding	19	5.55966e-6	9.52164e-4	3.2900
Egr1 $^{+/-}$	Growth Factor Binding	51	6.61498e-4	4.22062e-2	2.1944

[1] *GO Molecular Function* has 3,254 terms covering 15,499 (77%) of all 20,221 genes. 3,254 ontology terms were tested (100%) using an annotation count range of [1, Inf].

[2] With the WT genotype, the test set of 520 genomic regions picked 764 genes (4%) of all 20,221 genes. With the *Egr1*$^{+/-}$ genotype, the test set of 800 genomic regions picked 1,169 genes (6%) of all 20,221 genes.

Supplemental Table S3. GREAT analysis of the MSigDB Predicted Promoter Motif ontology [1] reveals an enrichment of genes that have *EGR* binding sites in WT mice.

Geno-type	Motif	Predicted Promoter	Rank	Raw P-Value	FDR Q-Val	Fold Enrichment
WT	NDDNNCACGTGNNNNN	*ARNT*	20	3.00960e-6	9.25452e-5	2.9139
WT	VGTGACGTMACN	*ATF2*	15	5.57507e-7	2.28578e-5	2.7737
WT	NNSATGAGTCATGNT	*BACH1*	29	1.36414e-5	2.89291e-4	2.6021
WT	NNRTGCAATMCCC	*DDIT3*	17	7.28236e-7	2.63450e-5	2.4324
WT	**WTGCGTGGGCGK**	***EGR1***	**27**	**1.07675e-5**	**2.45260e-4**	**2.4789**
WT	**NTGCGTRGGCGK**	***EGR2***	**1**	**2.13321e-18**	**1.31192e-15**	**5.4541**
WT	**NTGCGTGGGCGK**	***EGR3***	**13**	**4.61749e-7**	**2.18443e-5**	**5.0992**
WT	**WTGCGTGGGYGG**	***EGR4***	**12**	**2.74924e-7**	**1.40898e-5**	**2.8108**
WT	ACWTCCK	*ETV4*	19	1.42172e-6	4.60187e-5	3.2285
WT	SNNNCCNCAGGCN	*GTF3A*	24	7.88678e-6	2.02099e-4	2.6873
WT	BNCRSTTTCANTTYY	*IRF1*	30	1.46674e-5	3.00682e-4	2.3527
WT	NNANCACGTGNTNN	*MAX*	18	1.11254e-6	3.80120e-5	3.0033
WT	TGACAGKTTTAYGA	*MEIS1*	31	2.18218e-5	4.32916e-4	2.5874
WT	GCCAYGYGSN	*MYC*	2	7.34784e-11	2.25946e-8	3.7075
WT	NNNNNNNCACGTGNNNNN	*MYC*	23	6.10703e-6	1.63297e-4	2.9505
WT	NNTTGGCNNNNNNCCNNN	*NF1*	25	9.58678e-6	2.35835e-4	2.3658
WT	NTGGNNNNNNGCCAANN	*NF1*	26	1.03194e-5	2.44093e-4	2.3569
WT	AAANWWTGC	unknown	10	1.88094e-7	1.15678e-5	2.4642
WT	AATWTTCAACAG	unknown	14	4.77276e-7	2.09661e-5	2.4759
WT	WNWCACCTGWNN	*TCF8*	3	8.26277e-11	1.69387e-8	3.0925
Egr1$^{+/-}$	NNSATGAGTCATGNT	*BACH1*	28	5.20219e-5	1.14262e-3	2.1467
Egr1$^{+/-}$	ACWTCCK	*ETV4*	5	1.21166e-9	1.49035e-7	3.2847
Egr1$^{+/-}$	NNTKACGTCANNNS	*CREB1*	29	6.50405e-5	1.37931e-3	2.2516
Egr1$^{+/-}$	ANNCACTTCCTG	*ETS1*	23	1.17355e-5	3.13796e-4	2.3557
Egr1$^{+/-}$	MGGAAGTG	*GABPA*	2	6.64745e-11	2.04409e-8	2.3138
Egr1$^{+/-}$	BNCRSTTTCANTTYY	*IRF1*	10	1.74813e-7	1.07510e-5	2.3185
Egr1$^{+/-}$	GCCAYGYGSN	*MYC*	1	9.20735e-15	5.66252e-12	3.5803
Egr1$^{+/-}$	NGGGACTTTCCA	unknown	8	1.03914e-8	7.98841e-7	2.7315
Egr1$^{+/-}$	NNNNKGGRAANTCCCN	unknown	11	2.25198e-7	1.25906e-5	2.6142
Egr1$^{+/-}$	GCCNNNWTAAR	unknown	14	1.16542e-6	5.11952e-5	2.5523
Egr1$^{+/-}$	NGGGGAMTTTCCNN	unknown	15	1.34452e-6	5.51254e-5	2.2968
Egr1$^{+/-}$	SYATTGTG	unknown	22	8.68204e-6	2.42702e-4	2.1465
Egr1$^{+/-}$	GGGAMTTYCC	*RELA*	16	1.98324e-6	7.62309e-5	2.4143
Egr1$^{+/-}$	GGGRATTTCC	*RELA*	18	2.59623e-6	8.87046e-5	2.4206
Egr1$^{+/-}$	CCAWWNAAGG	*SRF*	20	6.86589e-6	2.11126e-4	3.1311
Egr1$^{+/-}$	NNTTTCCN	*STAT1*	3	1.01299e-10	2.07662e-8	3.1038
Egr1$^{+/-}$	CAGTTTCWCTTTYCC	*STAT1,2*	13	5.13684e-7	2.43012e-5	2.3034
Egr1$^{+/-}$	RNCAGCTGC	*TFAP4*	21	7.37284e-6	2.15919e-4	2.1150
Egr1$^{+/-}$	CAGCTGS	*UBP1*	6	3.29180e-9	3.37410e-7	2.8038
Egr1$^{+/-}$	SGGRNTTTCC	*REL*	9	1.54373e-8	1.05488e-6	2.5851

[1] *MSigDB Predicted Promoter Motifs* has 615 terms covering 8,450 (42%) of all 20,221 genes, and 120,192 term - gene associations.615 ontology terms (100%) were tested using an annotation count range of [1, Inf].

Supplemental Table S4 List of genes in WT mice with MSigDB Predicted Promoter Motifs matching *Egr1*, *Egr2*, *Egr3* and *Egr 4*.

			Relative gene expression of del(5q) vs. non-del(5q) MN[2]		
Gene	**MSigDB Predicted Promoter motif matches**	**EGR1 binding[1]**	**UC-1** (GSE39991)	**UC-2** (SRA061655)	**TCGA**
Adss	*Egr1, Egr4*	yes	1.03	0.89	1.00
Ap1g1	*Egr1, Egr4*	yes	1.47	0.89	0.90
Bcl6	*Egr3, Egr 4*	yes	**2.3**	1.18	0.89
Egr1	*Egr1, Egr2, Egr3, Egr4*	yes	**0.41**	0.71	**0.28**
Etf1	*Egr4*	yes	**0.66**	**0.46**	**0.57**
Gnai2	*Egr2, Egr4*	yes	1.12	1.11	1.04
Grb2	*Egr1, Egr2, Egr4*	yes	1.11	0.74	0.82
Hhex	*Egr1*	yes	0.72	0.95	0.90
Hyal2	*Egr1, Egr2, Egr4*	yes	0.93	1.38	1.00
Itpr1	*Egr2*	yes	**0.39**	0.66	0.92
Klf16	*Egr1*	yes	0.96	0.86	0.94
Klf3	*Egr1, Egr4*	yes	**2.8**	1.47	1.14
Lef1	*Egr1, Egr4*	yes	1.6	**14.81**	**2.16**
Mef2c	*Egr1*	yes	0.65	0.6	1.16
Myb	*Egr1, Egr2, Egr3, Egr4*	yes	0.57	1.11	1.02
Nrgn	*Egr1, Egr2, Egr4*	yes	**2.15**	1.39	**1.97**
Rnf24	*Egr1*	yes	1.74	1.16	1.05
Tcfe3	*Egr1*	yes	1.31	1.1	0.99
Ubtf	*Egr1*	yes	1.16	1.27	1.12
Atp1b2	*Egr1, Egr4*	No	2.14	11.38	5.89
Atp6v1c1	*Egr2, Egr3*	No	0.95	1.24	0.94
Coro1c	*Egr1, Egr 4*	No	1.67	0.99	1.10
Egr3	*Egr1, Egr2, Egr3, Egr4*	No	0.95	0.71	1.11
Ephb3	*Egr2*	No	1.06	0.35	1.09
Gnl1	*Egr2, Egr4*	No	1.04	1.27	1.06
Lrrc4	*Egr4*	No	1.5	1.59	1.07
Ntn1	*Egr4*	No	1.04	0.52	0.73
Pdgfb	*Egr1, Egr2, Egr3, Egr4*	No	1.11	0.98	
Ralgps2	*Egr1, Egr2, Egr3, Egr4*	No	1.05	0.27	0.76
Smad1	*Egr2*	No	0.73	3.08	1.06
Sox4	*Egr2*	No	0.25	1.41	0.80
Trib1	*Egr1, Egr4*	No	1.26	0.61	0.93
Zhx2	*Egr2*	No	0.64	1.51	1.16

[1] Because of the redundancy of EGR family binding sequences, we examined whether EGR1 binds to these genes by searching the ENCODE EGR1-specific ChIP dataset.

[2] Of the ChIP confirmed EGR1 target genes, gene expression, with a significant (Wilcox $P < 0.05$) 2-fold increase or 2-fold decrease (i.e.≤ 0.5) in at least one of the three data sets, is in bold.