www.sciencetranslationalmedicine.org/cgi/content/full/4/133/133ra57/DC1

Supplementary Materials for

Extended Survival of Glioblastoma Patients After Chemoprotective HSC Gene Therapy

Jennifer E. Adair, Brian C. Beard, Grant D. Trobridge, Tobias Neff, Jason K. Rockhill, Daniel L. Silbergeld, Maciej M. Mrugala, Hans-Peter Kiem*

*To whom correspondence should be addressed. E-mail: hkiem@fhcrc.org

Published 9 May 2012, *Sci. Transl. Med.* **4**, 133ra57 (2012)
DOI: 10.1126/scitranslmed.3003425

The PDF file includes:

SUPPLEMENTARY MATERIALS

SUPPLEMENTARY RESULTS

Extended RIS Studies: Confirmation of Clonal Predominance, Analysis of Common Integration Sites (CISs), and Over-Represented Gene Classes Near RIS

To confirm that identified dominant clones were truly enriched in the original WBC DNA pool and not the product of biased PCR amplification during processing for RIS analysis, we evaluated the number of identified shear sites associated with specific clonal sequences after mapping RIS to genomic locations (fig. S3A-C) (1). Using this analysis, both *PRDM16* (P1) and *HMGA2* (P2) clones were confirmed as valid dominant clones with multiple shear sites represented in mapped sequences.

For common integration site (CIS) analysis, RIS were grouped into those isolated after transplant but prior to chemotherapy (Chx 0), after 1-2 cycles of chemotherapy (Chx 1-2), and after ≥ 3 cycles of chemotherapy (Chx ≥ 3) (table S4). CIS analysis revealed that RIS clustering increased over time and with cycles of chemotherapy. The two clusters that contained the RIS in intronic regions of *PRDM16* and *HMGA2* that lead to clonal dominance were represented. Overall RIS clustering was only evident in groups Chx 1-2 and Chx ≥ 3, and there was no RIS clustering containing five unique RIS identified in windows <300kb in group Chx 0.

For over-represented gene class analysis, RIS were grouped as above (table S5). The gene ontology group “phosphoprotein” was the top ranked over-representation gene class in all RIS groups, and “acetylation” and “nucleus” were identified as over-represented gene classes in RIS groups Chx 1-2 and Chx ≥ 3. The gene class “proto-oncogene” was identified as over-represented in Chx ≥ 3 (table S5).

SUPPLEMENTARY DISCUSSION

Individual Patient Descriptions

P1 exhibited rapid recovery from the first two cycles of O^6BG/TMZ. Prior to receiving a third cycle, including an escalated dose of TMZ (590 mg/m^2), MRI imaging showed an increase in the size of the enhancing lesion with increased T2/FLAIR signal and the patient underwent a second surgical resection. MRI findings and tumor pathology were reviewed by the Institutional Data and Safety Monitoring Board (DSMB) and were felt to be consistent with pseudoprogression (2,3). Following a subsequent dose-escalated cycle, Grade IV thrombocytopenia and afebrile neutropenia were observed, but with prompt recovery to retreatment thresholds. P1 remains alive with no clinical evidence for disease progression at 33 months since diagnosis.

P2 received the highest cell dose and displayed the most rapid recovery following BCNU administration and transplant, with 50% gene marking in WBCs. Following administration of O^6BG/TMZ, P2 displayed modest myelosuppression; however, despite prompt PLT recovery, ANC remained just below retreatment criteria for up to 42 days after treatment during cycles 1 and 2. Despite this delay, P2 displayed a two-fold increase in the number of circulating gene modified CD34$^+$ CFCs, suggesting *in vivo* selection of gene-modified cells. Following cycle 3 on day 124 post-transplant, P2 displayed excellent chemoprotection of both neutrophil and platelet compartments with rapid recovery, correlating with the time after transplantation for long-term repopulating cells to assume hematopoiesis mentioned above. Given sub-total initial resection and unmethylated tumor MGMT promoter status, both predicted to reduce progression-free survival, 9 months of progression-free survival is notable. Furthermore, radiation therapy was delayed (7 weeks after resection) in P2, which correlates with earlier time to progression. P2 succumbed to disease progression at 14 months after transplant. At the most recent time point

prior to death, P2 displayed a provirus copy number of 0.64 in granulocytes (64% gene-modified cells assuming one provirus copy per cell).

P3 received the lowest cell dose and displayed the slowest engraftment following BCNU administration and transplant. Despite this, P3 demonstrated the most robust protection of the neutrophil compartment during O^6BG/TMZ chemotherapy, with ANC recovery to retreatment thresholds within 28 days of each prior cycle. In contrast, platelet recovery, though independent of transfusion or other support, was significantly delayed, with the longest delay being 85 days between cycles 2 and 3. We did attempt to evaluate gene marking in the platelet progenitor (megakaryocyte; MK) population in CFU assays in which $CD34^+$ BM cells were cultured in conditions skewed toward MK-CFU development including growth factors thrombopoietin (50ng/mL), interleukin-3, interleukin-6 and stem cell factor (10 ng/mL each). When individual colonies grown under these conditions were cytospun and subjected to immunostaining with an antibody directed against GpIIb, a marker of MK progenitors, we observed at least partial GpIIb, in all colonies stained. When gene marking in colonies grown under the same conditions was assessed, the percent of marked CFUs (84.6%; 11 out of 13 colonies assessed) did not differ significantly from CFU marking observed in colonies grown under standard colony assay conditions described in *Methods* (77%). However, colonies picked can either be assessed for gene marking or for GpIIb immunostaining, but not both, making confirmation of MK-like CFU gene marking inferential in these studies. Real-time PCR analysis of bulk BM WBCs at this time point indicated a provirus copy number of 57.4, which corresponds to 57.4% of cells harboring transgene-containing provirus if we assume one copy per genome. This correlates with the number of PB WBCs harboring provirus (52%), assessed by the same assay within one week of the BM assay. However, presence of provirus backbone alone does not rule out subset-specific transgene silencing. We have attempted to determine the presence of MGMT transcript in the MK cell population ($CD34^-/CD41a^+$) in this patient, but have been unsuccessful in obtaining a high purity cell population from which sufficient RNA to conduct such an analysis would be

possible. P3 succumbed to disease progression at 20 months following transplant, displaying a provirus copy number of 0.61 in granulocytes (61% gene-modified cells assuming one provirus copy per cell) at the most recent time point prior to death.

Analysis of Common Integration Site (CIS) and Over-Represented Gene Classes near Retrovirus Integration Site (RIS)

CIS analysis is important to identify genomic loci that are favored for retrovirus integration and/or how provirus integration influences cellular processes. While no CIS were identified after transplant and prior to any chemotherapy, Chx 0 (narrowest window was >300kb), CIS were identified as a function of time and chemotherapy (Chx 1-2 and Chx ≥ 3) as identified in previous clinical trials in the absence of chemotherapy {33040}. Both dominant clone loci, PRDM16 and HMGA2, were identified as CISs in RIS groups Chx 1-2 and Chx ≥ 3 (table S2). RIS clustering at both PRDM16 and HMGA2 are particularly interesting based on the expansion results from P1 and P3, so we have also included a schematic figure (fig S2) to more thoroughly describe the profiles around these sites. Varying genotoxic effects of gammaretrovirus vectors are well-described, but this level of analysis in this unique *in vivo* selection setting adds important biological information regarding the use of gammaretrovirus vectors and can be used to intelligently develop safer strategies of gene modification and at the same time provide valuable information regarding potentially HSC expanding genes or microRNA regulatory domains.

Over-represented gene classes, gene ontology groups, near RIS can provide insight into cellular processes that when activated by provirus integration lead to an expansion and subsequent over-representation of clonal contribution. The most over-represented gene ontology groups across the RIS groups described above are “phosphoprotein”, “acetylation” and “nucleus”. The gene ontology group “proto-oncogene” was identified as over-represented in Chx ≥ 3, but as the least significant of the gene classes over-represented for that RIS group. The

proto-oncogene gene ontology group was not previously identified in our pre-clinical studies {38473, 36077}, but the level of analysis of RIS populations we are now able to achieve with MGS-PCR that we have developed warrants reexamination of previous studies. Collectively, CIS analysis and gene ontology enrichment provide valuable data regarding the safety of a retrovirus-based approach and also direct future investigation into previously unidentified genes and regulatory elements critical in controlling normal hematopoiesis.

SUPPLEMENTARY MATERIALS AND METHODS

Gammaretrovirus Vector and Detailed Transduction Protocol

The gammaretrovirus vector used in these studies MND-GRS-P140K has been previously described (4). Virus titers used in these studies ranged from 6.3×10^4 to 1.9×10^5 infectious units per mL (IU/mL) of virus-conditioned media.

Following CD34 enrichment, at least 2.5×10^6 $CD34^+$ cells/kg for each patient were cultured for a minimum of 30 hours in StemSpan SFEM media (Stem Cell Technologies, Vancouver, BC) containing 100ng/mL of the following growth factors: recombinant human G-CSF (Neupogen, Amgen, Thousand Oaks, CA), recombinant human stem cell factor (SCF; Cellgenix, Antioch, IL), recombinant human Flt3-ligand (Cellgenix), and recombinant human thrombopoietin (TPO; Cellgenix). Cells were then harvested and replated in viral conditioned media (DMEM) containing the same growth factor cocktail mentioned above, as well as 4µg/mL protamine sulfate (Abraxis Pharmaceuticals, Schaumburg, IL) at a cell density of $0.5\text{-}1.0 \times 10^6$ cells/mL on culture vessels pre-coated with fibronectin fragment CH-296 (Retronectin®; Takara Biomedical, Japan). Prior to cell plating, CH-296-coated vessels were pre-loaded 2x with viral conditioned media for a minimum of 15 minutes at room temperature. Transduction in viral-conditioned media was allowed to proceed under incubation for a period of 4 hours. After which, cells were incubated on CH-296-pre-coated vessels for a minimum of 12 hours in StemSpam SFEM containing the same growth factor cocktail mentioned above. Following this incubation period a second, 4-hour transduction was conducted as before, but without pre-loading CH-296-coated vessels with viral conditioned media. After the second transduction period, cells were harvested, washed in PlasmaLyte-A (Baxter, Deerfield, IL) and prepared for infusion by resuspension in PlasmaLyte-A containing 20% human serum albumin (HSA; ZLB Behring, Kankakee, IL). Release criteria for infusion included a negative gram stain, ≥ 90% viability by

trypan blue staining and an endotoxin level determination. All incubation periods were performed at 37℃ with 85% relative humidity and 5% CO_2.

Modified genome sequencing (MGS)-PCR for detection of retrovirus integration sites. High molecular weight DNA was extracted from WBC samples derived from either PB, immune-purified PB subsets or BM using the Qiagen Gentra Puregene Blood DNA Isolation Kit (Qiagen, Valencia, CA), as per the manufacturer's protocol, from no less than 1.5×10^7 starting cells. Extracted DNA was then quantitated on a NanoDrop spectrophotometer and 0.3-3 µg total DNA was subjected to subsequent RIS analysis.

Three hundred nanograms of the same DNA sample from each time point in each patient was also subjected to quantitative real-time PCR (Taqman) specific for proviral sequences (specifically the MND 5-prime non-translated region to determine the level of gene-modified cells). In these reactions, which are performed in replicates of 2-3 reactions per sample, signal was normalized to a β-globin standard (control for DNA integrity and concentration) and proviral copy number was determined by standard curve method using DNA from a baboon B-cell line which has been verified to contain a single provirus integrant per genome. This analysis provides a representative percent of gene-modified cells the PB by multiplying the average proviral copy number obtained by 100% (e.g. If a given sample yields an average proviral copy number of 0.12, this corresponds to 12% assuming that each gene-modified cell contains one provirus copy).

For RIS analysis, DNA was sheared by nebulization using pressurized nitrogen to avoid use of restriction enzymes. This shearing technique results in a range of DNA fragment lengths, which are then resolved by agarose gel electrophoresis against a 1kb DNA ladder. Fragments ≥600bp are excised and purified using the Qiagen Qiaquick Gel Purification kit according to the manufacturer's protocol. Purified fragments were then subjected to polishing and modified linkers were ligated to the fragments using the 454/Roche-GS 20 DNA Library Preparation Kit

as per the manufacturer's protocol. The samples were then subjected to two sequential, nested, exponential PCR reactions in order to amplify the vector-genome junctions present. The first of these two PCR reactions incorporated a biotin-tagged long-terminal repeat (LTR-)-specific primer allowing for further purification of amplified fragments by streptavidin-coupled bead capture using Invitrogen's DynaBeads M-280 according to the manufacturer's protocol. Purified products from the first nested PCR were then subjected to a second nested, exponential PCR which incorporated 1 of 100 different multiplex identifier tags (MID1-100; supplied by Roche) in the LTR-specific primer to allow for multiplexing of different time points within the same patient for longitudinal analyses. Following the second nested PCR reaction, samples were purified to remove short fragments and submitted for sequencing on the 454/Roche Titanium system.

To be qualified as a unique integration site sequence (clone), an LTR junction must be identified in the amplified fragment and the adjacent sequence must be an appropriate length (at least 19 nucleotides) to definitively align to a single locus within the human genome. This known LTR sequence identification and chromosomal alignment requirement make this assay self-validating. Once these unique integration sites were identified, a "capture frequency" for each site was determined by the number of times that a specific integration site was sequenced relative to the total number of sequenced sites identified in the same sample. Using these methods, any over-represented clone of interest should be detectable at any given time point and sequence information can provide the platform necessary for further tracking of over-represented clones.

PCR-based Clone Tracking

WBC DNA was extracted using either the Qiagen Gentra Puregene Cell Kit or the Qiagen QIAamp DNA Blood Mini Kit (Qiagen, Valencia, CA) per the manufacturer's protocol. Extracted DNA was then quantitated on a NanoDrop spectrophotometer and samples were subjected to clone specific PCR as previously described {36077} using the following primer sets: PRDM16

Clone 3087539 forward 5'-CAA ACC TAC AGG TGG GGT CT-3', reverse 5'-GGG TTA AAT GCT GAG CAA GC-3'; PRDM16 Clone 2988812 first primer set: forward 5'-ATC CTG TTT GGC CCA TAT TC-3', reverse 5'-GCT CAC AGG ACA AGG AGG AC-3'; second, nested PCR reaction primer set: forward 5'-CAA ACC TAC AGG TGG GGT CT-3', reverse 5'-AGG GGC TAG ATG GGA CTT GT-3'; HMGA2 Clone 3087539 first primer set forward 5'-ATC CTG TTT GGC CCA TAT TC-3', reverse 5'-AAA TAC CAC CCC AAC CCA CT-3'; second nested PCR reaction primer set: forward 5'-CAA ACC TAC AGG TGG GGT CT-3', reverse 5'-CAC TTG TCA GCC TCA GAG CA-3'; the MND provirus backbone forward 5'-ACA CAG AAC GAT GCT GCA GC-3', reverse 5'-TTG CGC CTG CGT CTG TAC TA-3'; β-Actin forward 5'-TCC TGT GGC ATC GAC GAA ACT-3', reverse 5'-GAA GCA TTT GCG GTG GAC GAT-3'. Resulting PCR products were resolved in 2% agarose by gel electrophoresis and visualized by ethidium bromide staining under ultraviolet light.

Transcript Analysis

WBC RNA was extracted using either RNA Stat-60 (Amsbio, Lake Forest, CA) or the Qiagen RNeasy Mini Kit per the manufacturer's protocols. Extracted RNA was quantitated on a NanoDrop spectrophotometer and samples were subjected to reverse transcription into cDNA using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Life Technologies, Carlsbad, CA) according to the manufacturer's protocol. Resulting cDNA was quantitated on a NanoDrop spectrophotometer and samples were subjected to PCR to amplify either PRDM16 transcript cDNAs (Applied Biosystems pre-optimized gene expression kit #Hs00922673_m1), HMGA2 transcript cDNAs (Applied Biosystems pre-optimized gene expression kit #Hs00971725_m1), MGMT transcript cDNAs (forward 5'-CAC CAG ACA GGT GTT ATG G-3', reverse 5'-GCT GCA GAC CAC TCT GTG-3') or GAPDH transcript cDNAs (Applied Biosystems pre-optimized gene expression kit #Hs03929097_g1). Thermalcycler conditions for these reactions consisted of 50℃ fo r 2 minutes, 95℃ for 10 minutes, followed by

41 cycles of 95℃ for 15 seconds and 60℃ for 1 min utes, then a final hold at 4℃. Resulting PCR products were resolved in 2% agarose by gel electrophoresis and visualized by ethidium bromide staining under ultraviolet light.

Analysis of over-represented genes near RIS as a function of time and chemotherapy regimens

The closest gene was defined by the Refseq gene with the transcription start site (TSS) closest to the RIS and the gene ontology analysis was carried our as previously described (5) except the window size surrounding the TSS was restricted to 2.5kb.

Analysis of Common Integration Sites (CISs)

To analyze RIS clustering the smallest window size that contained five unique RIS were identified and ranked as previously described (6).

Intracellular MGMT staining of hematopoietic cells

Intranuclear staining for MGMT expression was assessed using the Cytomation IntraStain Fixation and Permeabilization Kit (Dako, Carpinteria, CA) according to manufacturers recommendations. Antibodies included monoclonal mouse anti-human MGMT (Kamiya) or Rabbit anti-mouse IgG1 (Dako), and phycoerythrin-conjugated (PE) anti-mouse IgG_1 secondary antibody (Dako). PE fluorescence was assessed by flow cytometry on a FACS Vantage or FACS Calibur (Becton Dickinson, San Jose, CA). At least 20,000 events (forward and right-angle light scatter-gated) were assessed per sample. Three negative controls were used: (1) an unstained sample of the transduced cells from study patients, (2) and isotype stained sample of the transduced cells from study patients, and (3) an MGMT-stained sample from a normal, non-transplanted human patient. Flow-cytometric data were analyzed by FlowJo v8.5.2 to v8.8.6 or

CELLQuest v3.1f software (Becton Dickinson) with gating to exclude fewer than 0.1% control cells in the relevant region.

Pharmacokinetic studies

Blood was collected in lithium heparin tubes before, and at 0, 0.25, 0.5, 1, 1.5, 2, 4 and 6 hours after TMZ administration to assess disposition of TMZ. Samples were processed for plasma by centrifugation within 5 minutes of being drawn, phosphoric acid was added to a final concentration of 0.5% and samples were stored at -80℃ until analysis. Plasma levels of TMZ were measured by isocratic high-performace liquid chromatography (7) and pharmacokinetic estimates were generated using WinNonlin software (PharSight Corp., Mt. View, CA).

Patient clinical status following transplant and subsequent combination chemotherapy.

As indicated in table S1, all three patients were between the ages of 51 and 56 years at the time of diagnosis. P1 and P3 were reported to have gross total initial surgical resections, while P2 received a subtotal, but >50% resection of the original tumor mass. All patients demonstrated unmethylated MGMT promoters, indicating high expression of wild-type MGMT, in resected tumor tissue. P1 and P2 demonstrated stable disease following radiotherapy, while P3 demonstrated a recurrence of the primary lesion with appearance of a new distal lesion following radiotherapy. Using post-radiotherapy MRI scans as baseline, all three patients demonstrated stable disease by MRI through G-CSF mobilization and collection procedures and BCNU administration, as well as through at least 2 cycles of O^6BG/TMZ chemotherapy. Follow-up for all three patients is now at least 18 months, and disease evaluation at 12 months showed 2 of the 3 patients with stable disease. MRI studies for P1 and P3 are shown in fig. S5. All patients were neurologically well (Karnofsky Performance Scores (KPS) ≥ 90%) and alive at 12 months since diagnosis, the reported median survival for this disease subgroup.

SUPPLEMENTARY REFERENCES

Reference List as of 02-10-2012

1. C. C. Berry, N. A. Gillett, A. Melamed, N. Gormley, C. R. Bangham, F. Bushman, Estimating abundances of retroviral insertion sites form DNA fragment length data. *Bioinformatics* prepublished online January 11, 2012; doi:10.1093/bioinformatics/bts004 (in press).
2. A. A. Brandes, E. Franceschi, A. Tosoni, V. Blatt, A. Pession, G. Tallini, R. Bertorelle, S. Bartolini, F. Calbucci, A. Andreoli, G. Frezza, M. Leonardi, F. Spagnolli, M. Ermani, MGMT promoter methylation status can predict the incidence and outcome of pseudoprogression after concomitant radiochemotherapy in newly diagnosed glioblastoma patients. *J. Clin. Oncol.* **26,** 2192-2197 (2008).
3. A. A. Brandes, A. Tosoni, F. Spagnolli, G. Frezza, M. Leonardi, F. Calbucci, E. Franceschi, Disease progression or pseudoprogression after concomitant radiochemotherapy treatment: pitfalls in neurooncology (Review). *Neuro-oncol* **10,** 361-367 (2008).
4. B. C. Beard, G. D. Trobridge, C. Ironside, J. S. McCune, J. E. Adair, H.-P. Kiem, Efficient and stable MGMT-mediated selection of long-term repopulating stem cells in nonhuman primates. *J. Clin. Invest.* **120,** 2345-2354 (2010).
5. B. C. Beard, D. Dickerson, K. Beebe, C. Gooch, J. Fletcher, T. Okbinoglu, D. G. Miller, M. A. Jacobs, R. Kaul, H.-P. Kiem, G. D. Trobridge, Comparison of HIV-derived lentiviral and MLV-based gammaretroviral vector integration sites in primate repopulating cells. *Molecular Therapy* **15,** 1356-1365 (2007).
6. G. D. Trobridge, D. G. Miller, M. A. Jacobs, J. M. Allen, H.-P. Kiem, R. Kaul, D. W. Russell, Foamy virus vector integration sites in normal human cells. *PNAS* **103,** 1498-1503 (2006).

7. M. N. Kirstein, J. C. Panetta, A. Gajjar, G. Nair, L. C. Iacono, B. B. Freeman, III, C. F. Stewart, Development of a pharmacokinetic limited sampling model for temozolomide and its active metabolite MTIC. *Cancer Chemother. Pharmacol.* **55,** 433-438 (2005).

Supplemental Table S1. Individual Patient Data

Patient	Age at diagnosis	KPS at diagnosis (%)	Extent of initial surgical resection	Status at 12 months after diagnosis
P1	56	100	Gross total	Alive, SD[†]
P2	51	100	>90%	Alive, PD (day 262)
P3	52	90	Gross total	Alive, SD

SD: stable disease, PD: progressive disease, as measured by a ≥ 25% increase in the sum of the products of the cross-sectional diameters of the enhancing lesion. [†]True disease progression has not been documented in this patient to date; however, pseudoprogression was evident at 183 days since diagnosis.

SupplementalTable S2. Pharmacokinetics of TMZ

Patient	Dose (mg/m^2)	C_{max} (ug/mL)	T_{max} (h)	$T_{1/2}$ (h)	$AUC_{0\rightarrow\infty}$ (ug/mL x h)	CL ($L/h/m^2$)
P1	472	18.01	2.0	1.88	89.31	6.70
P2	472	21.66	1.5	2.12	70.43	5.28
P3	472	24.12	0.5	2.19	93.86	5.03

C_{max}: Maximum TMZ concentration observed in plasma

T_{max}: Time from administration until plasma C_{max} was observed

$T_{1/2}$: TMZ half-life in plasma as observed over the time course studied

$AUC_{0\rightarrow\infty}$: Area under the concentration-time curve

CL: Plasma clearance

Supplemental Table S3. Unique RISs in Glioblastoma Patients

Patient	DPT (Chx)	Unique RIS
1	28	441
2	31	345
3	39	225
Total RIS Before Chx		**1,011**
1	49-104 (1-2)	724
2	66-101 (1-2)	1,378
3$^{\alpha}$	76-153 (1-2)	2,576
Total RIS After 1-2 Cycles of Chx		**4,678**
1$^{\beta}$	172-509 (3-9)	490
2$^{\gamma}$	192-424 (3)	2,430
3$^{\delta}$	230-432 (3-4)	3,823
Total RIS After ≥ 3 Cycles of Chx		**6,743**
Total RIS Combined		**12,432**

Unique RIS were not counted multiple times. The unique RIS is included in the first group in which it was detected and then not counted in subsequent time points.

Chx: Cycles of O^6BG/TMZ

α : Includes RIS identified in BM at DPT132

β : Includes RIS identified in CD15 sorted granulocytes and BM at DPT 384, forward and side scatter sorted granulocytes and CD3 sorted lymphocytes at DPT 417, CD3 sorted lymphocytes at DPT 423, and BM at DPT 509

γ : Includes RIS identified in ficoll isolated granulocytes and CD3 sorted lymphocytes at DPT 424

δ : Includes RIS identified in forward and side scatter sorted granulocytes and CD3 sorted lymphocytes at DPT 396

Supplemental Table S4. CIS in Glioblastoma Patients

Chromosome	RIS	Window Size	CIS Name	Nearest RefSeq Gene
Chx 0 (n=1,012)				
NA				
Chx 1-2 (n=4,573)				
Chr22	28,181,158	794	CIS Chx1-2-1	MN1
Chr8	121,076,851	2,434	CIS Chx1-2-2	COL14A1
Chr18	21,441,723	2,549	CIS Chx1-2-3	TTC39C
Chr12	4,378,916	2,725	CIS Chx1-2-4	CCND2
Chr1	94,556,923	3,104	CIS Chx1-2-5	ABCA4
Chr22	37,677,576	3,126	CIS Chx1-2-6	CYTH4
Chr11	121,451,894	3,505	CIS Chx1-2-7	SORL1
Chr17	43,300,628	3,541	CIS Chx1-2-8	FMNL1
Chr19	18,131,505	3,800	CIS Chx1-2-9	ARRDC2
Chr4	141,017,227	4,655	CIS Chx1-2-10	MAML3
Chr16	88,856,441	4,677	CIS Chx1-2-11	FAM38A
Chr1	92,924,335	5,503	CIS Chx1-2-12	GFI1
Chr5	131,756,281	6,299	CIS Chx1-2-13	C5orf56
Chr11	1,912,377	6,383	CIS Chx1-2-14	TNNT3
Chr1	43,795,091	8,144	CIS Chx1-2-15	MPL
Chr12	***66,350,404***	***8,675***	***CIS Chx1-2-16***	RPSAP52
Chr22	23,557,544	8,744	CIS Chx1-2-17	BCR
Chr12	4,349,029	10,382	CIS Chx1-2-18	CCND2
Chr14	100,527,754	11,240	CIS Chx1-2-19	EVL
Chr7	142,160,255	11,456	CIS Chx1-2-20	LOC730441
Chx ≥3 (n=6,196)				
Chr22	28,181,775	312	CIS Chx≥3-1	MN1
Chr4	55,685,938	314	CIS Chx≥3-2	KIT
Chr19	13,840,303	1,032	CIS Chx≥3-3	CCDC130
Chr12	***66,357,139***	***1,100***	***CIS Chx≥3-4***	RPSAP52
Chr22	23,559,095	1,267	CIS Chx≥3-5	BCR
Chr12	4,380,856	1,600	CIS Chx≥3-6	CCND2
Chr17	16,301,187	1,619	CIS Chx≥3-7	UBB
Chr1	***2,987,998***	***1,922***	***CIS Chx≥3-8***	PRDM16
Chr17	1,991,732	1,952	CIS Chx≥3-9	HIC1

Chr18	42,279,530	2,031	CIS Chx≥3-10	SETBP1
Chr18	42,086,516	2,048	CIS Chx≥3-11	SETBP1
Chr1	2,761,302	2,182	CIS Chx≥3-12	ACTRT2
Chr3	169,026,369	2,187	CIS Chx≥3-13	MECOM
Chr1	***3,099,149***	***2,304***	***CIS Chx≥3-14***	PRDM16
Chr3	168,930,765	2,529	CIS Chx≥3-15	MECOM
Chr1	109,394,588	2,571	CIS Chx≥3-16	AKNAD1
Chr20	47,486,016	2,944	CIS Chx≥3-17	PREX1
Chr20	17,947,292	3,474	CIS Chx≥3-18	SNX5
Chr10	22,542,135	3,839	CIS Chx≥3-19	COMMD3
Chr21	48,041,103	3,966	CIS Chx≥3-20	PRMT2

Chx 0: no CIS with a window size <100kb was identified

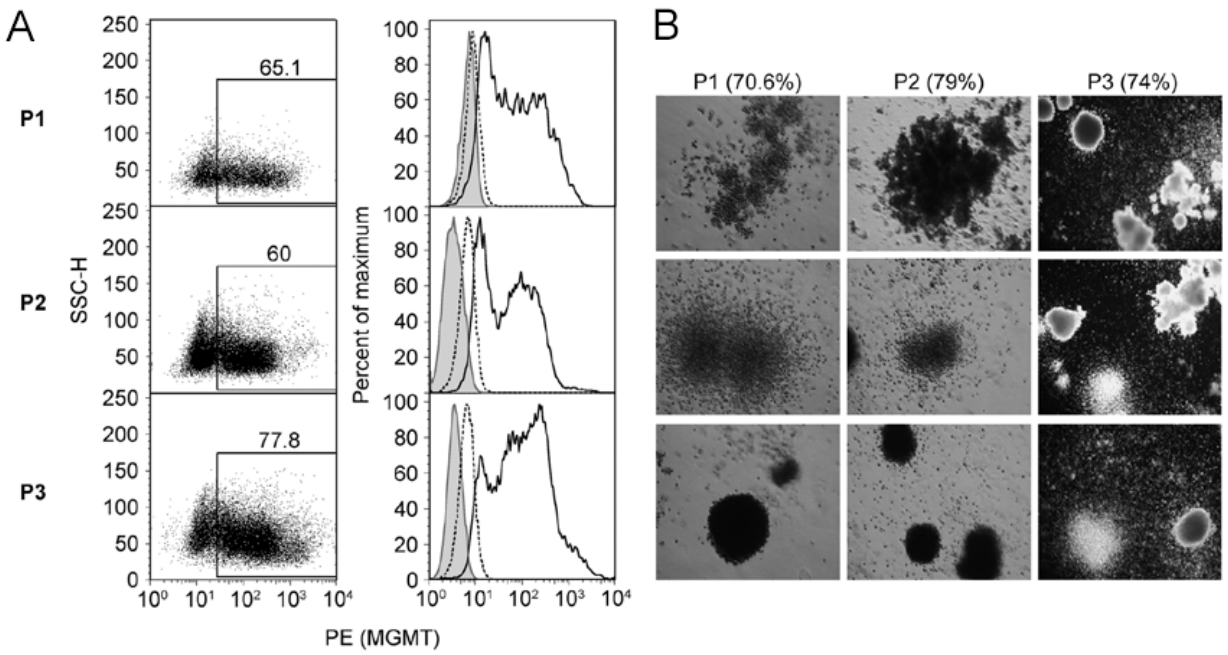
Bold italic represents *PRDM16* and *HMGA2* CIS

Chromosomal positions are listed using the RIS in the cluster with the lowest base pair position

Nearest RefSeq Gene = The closest RefSeq gene transcription start site

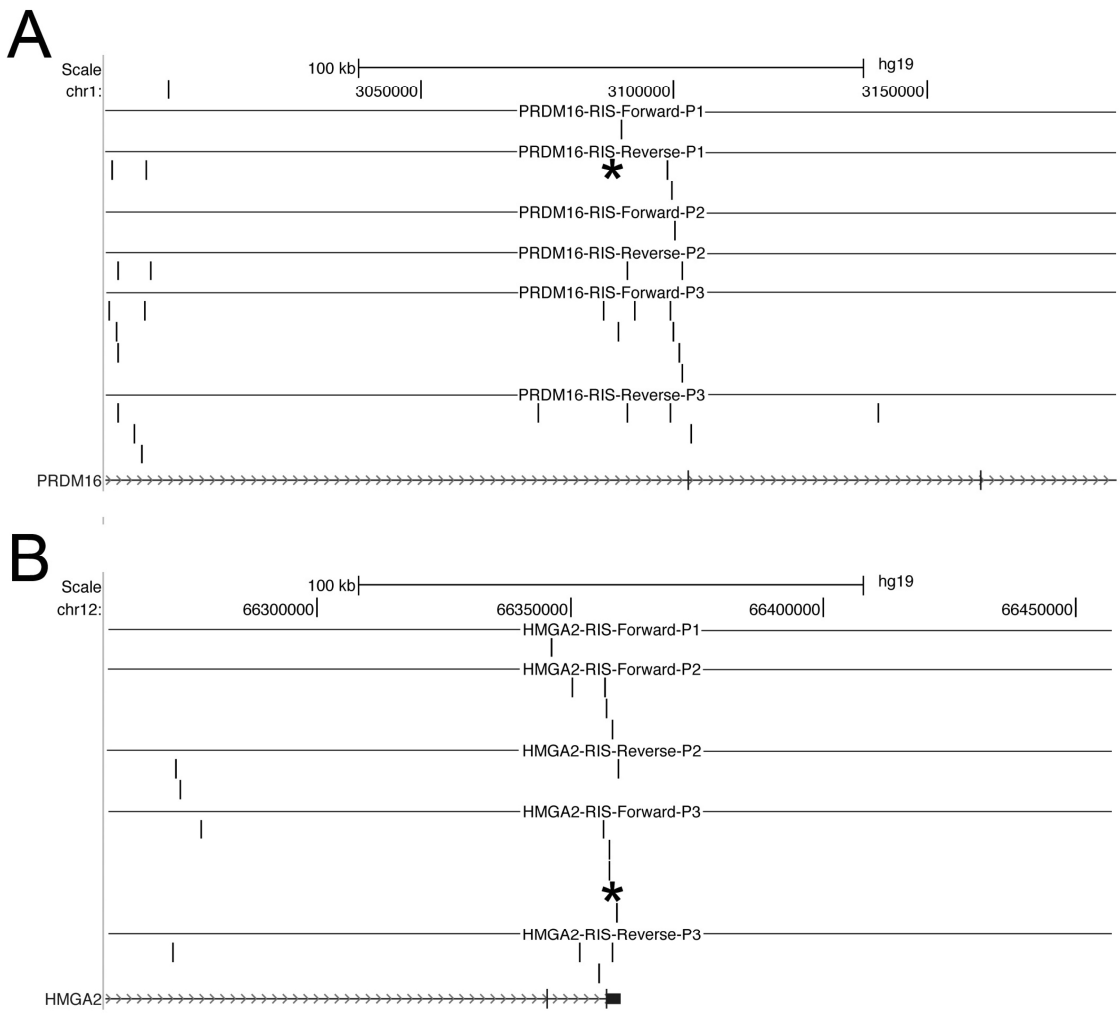
Supplemental Table S5. Over-Represented Gene Classes Near RIS

Keyword Term	# RIS Queried	# In Class	% In Class	Fisher	Bonferroni
Chx 0					
Phosphoprotein	349	167	47.85	7.45E-07	2.39E-04
Chx 1-2					
Phosphoprotein	1406	666	47.37	2.62E-22	1.55E-19
Nucleus	1406	384	27.31	2.42E-09	1.43E-06
Acetylation	1406	244	17.35	5.07E-07	2.99E-04
Apoptosis	1406	53	3.77	1.66E-06	9.78E-04
Chx ≥3					
Phosphoprotein	1822	812	44.57	1.08E-17	7.07E-15
Acetylation	1822	321	17.62	1.26E-09	8.26E-07
Nucleus	1822	478	26.23	1.14E-08	7.48E-06
Transcription Regulation	1822	248	13.61	1.09E-07	7.14E-05
Transcription	1822	249	13.67	4.31E-07	2.83E-04
Proto-oncogene	1822	43	2.36	6.00E-06	0.00393484



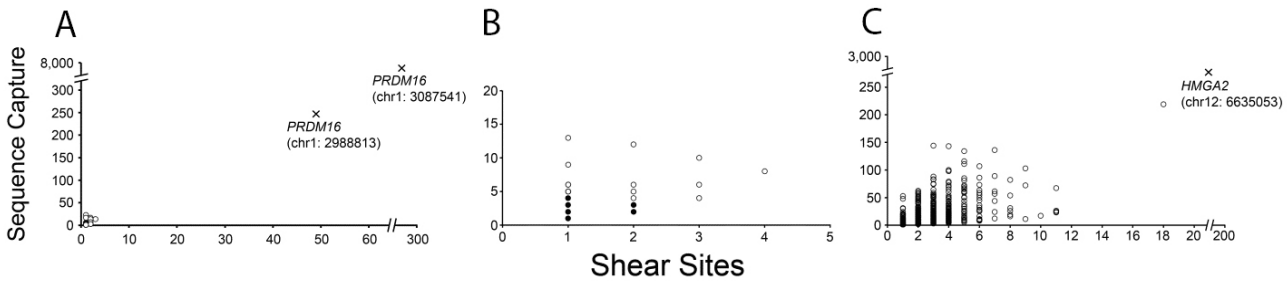
Supplemental Fig. S1. Efficient transduction of CD34$^+$ hematopoietic cells with a P140K-expressing PhGALV-pseudotyped gammaretrovirus.

(A) MGMT expression in P140K- transduced patient CD34$^+$ cells *in vitro* as assessed by intracellular MGMT staining analyzed by flow cytometry 72 hours following transduction. Dot plots represent gated forward scatter/side scatter cell populations graphed by side scatter (y-axis) and PE fluorescence intensity (x-axis). Histogram plots show unstained cell populations (dark gray peak), isotype-stained populations (dashed lines) and MGMT-stained populations (solid lines) for each patient as a percentage of the maximum number of cells gated (y-axis) and PE fluorescence intensity (x-axis). (**B**) Efficient colony formation from CD34$^+$ cells cultured for transduction. Photographs represent colonies derived from each patient. Individual colonies were isolated and evaluated for presence of provirus by PCR. Percentage of provirus$^+$ colonies is indicated in parenthesis next to patient identifiers (P1, P2 and P3).



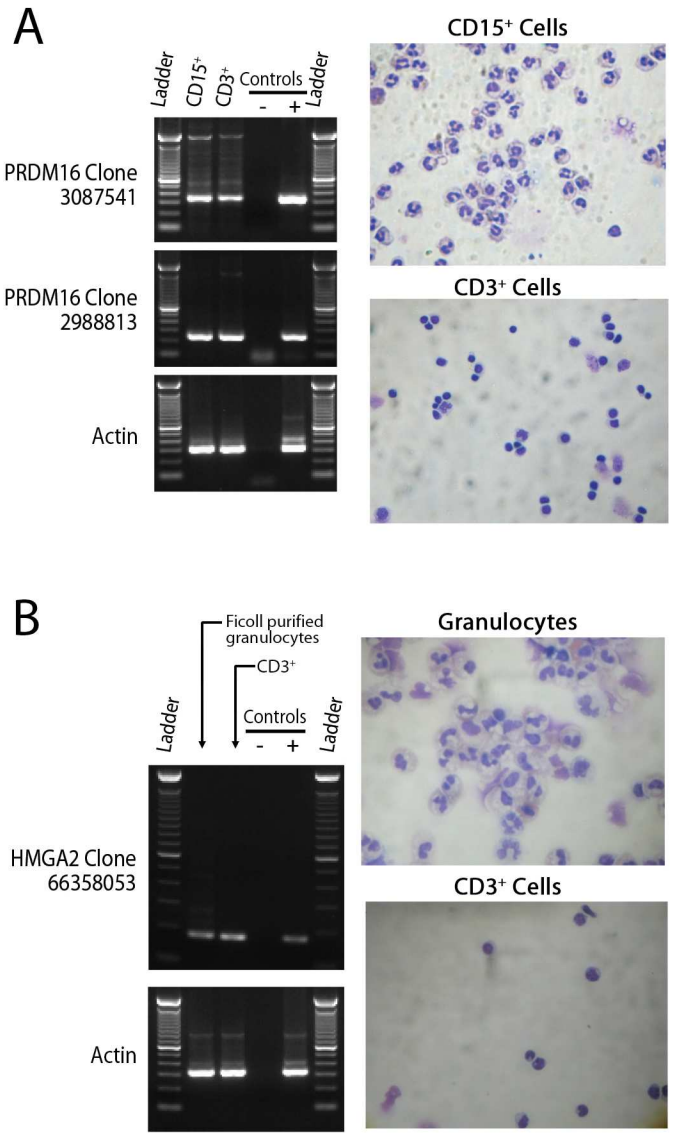
Supplemental Fig. S2. RIS distribution at the *PRDM16* and *HMGA2* loci.

Schematic representation of RIS distribution in GBM patients at the PRDM16 (**A**) and HMGA2 (**B**) loci. RISs identified in the dominant clones observed in P1 (PRDM16 chr1: 3807541) and P3 (HMGA2 chr12: 66358053) are indicated with asterisks. The genomic regions analyzed are 100kb upstream and downstream from the two indicated RISs. Each identified RIS is represented by a small, black vertical line. RISs are sorted by patient (P1, P2, and P3) the provirus orientation (Forward or Reverse). The Refseq gene schematics are included at the bottom of each panel. Introns are indicated by hashed arrowed lines, exons are indicated by black vertical lines and the 3` untranslated region for HMGA2 is indicated with a black rectangular box.



Supplemental Fig. S3. Retrovirus integration site sequence product analysis. Retrovirus integration sites (RIS) localized to a specific genomic location are plotted as a function of the absolute number of times that particular RIS was sequenced (Sequence Capture) versus the number of different products or lengths the “Sequence Capture” group was comprised of as a result of nebulization (Shear Sites). Closed circles in panel B represent >10 unique RIS with identical shear site and sequence capture values.

Supplemental Fig. S4



Supplemental Figure S4. Clonal tracking by PCR in mature WBC lineages. (A) Agarose gel electrophoresis of PCR products obtained from blood cell subset-specific DNA samples from P1 at day 472 after transplant, corresponding to each emergent *PRDM16* clone identified or Actin as a control for DNA integrity in the PCR reaction. Photographs show hematoxylin and eosin (H & E) stained cytospins of the cell population from which DNA was extracted for granulocyte and lymphocyte subsets specifically. **(B)** Agarose gel electrophoresis of PCR products obtained from P3 blood cell subset-specific DNA samples at day 396 after transplant, corresponding to the

prevalent *HMGA2* clone identified or Actin as a control for DNA integrity in the PCR reaction. Photographs show hematoxylin and eosin (H & E) stained cytospins of the cell population from which DNA was extracted for granulocyte and lymphocyte subsets specifically. Enriched subpopulations submitted for clone tracking were >98% purity by light microscopy visualization following H & E staining.