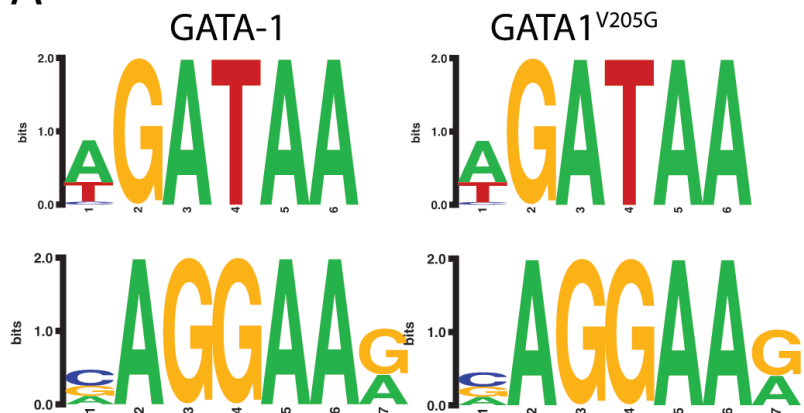
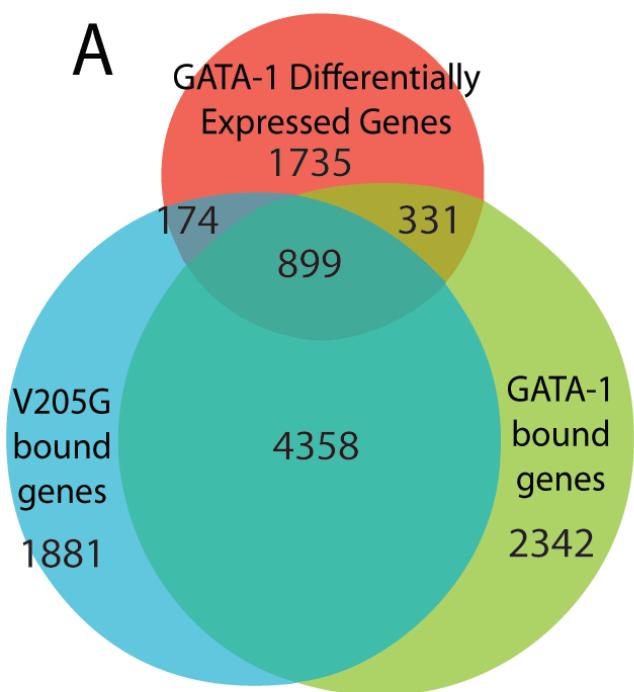


A



B

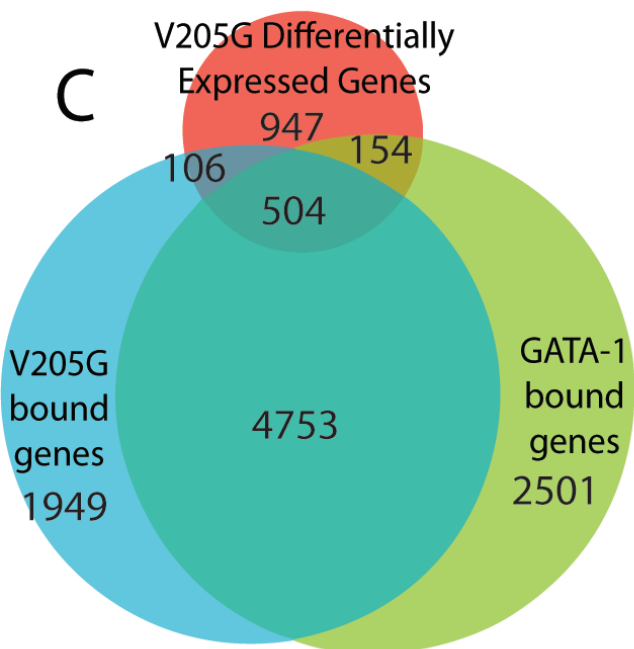
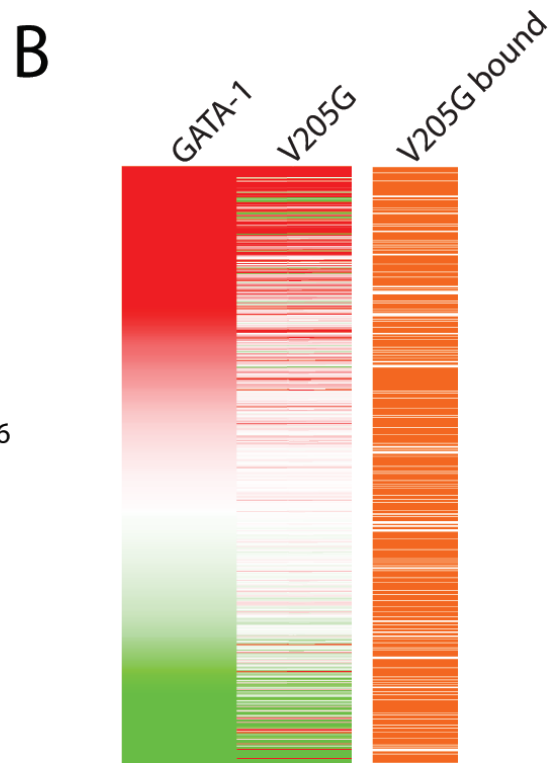
Output Motif #	GATA1 Only	Sites	E-value	Closest Match	V205G Only	Sites	E-value	Closest Match
1		661	1.5E-179	GATA-1		1232	3.6E-494	GATA-1
2		677	8.4E-160	GATA1		1232	5.8E-447	GATA-1
3		617	6.8E-115	FEV (ETS)		233	2.6E-41	Tcfap2e (AP-2)
4		713	2.5E-114	FEV (ETS)		549	1.6E-27	Myf6
5		497	4.6E-63	Myf (E-box)		62	1.7E-12	Zfp58
6		139	9.3E-22	KLF4 (CACCC)		124	1.1E-7	Zfp58



GATA-1

		Diff. expressed	
		No	Yes
Bound	No	21033	1909
	Yes	6682	1230

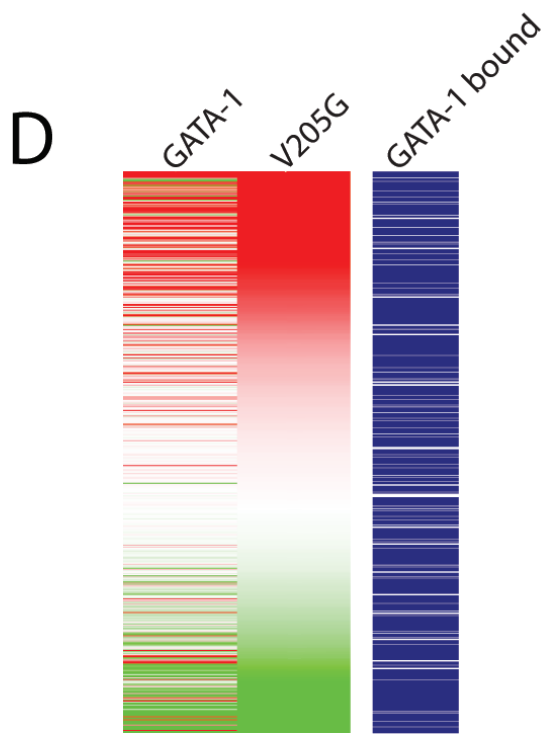
χ -squared= 335.26; p-value<2.2E-16



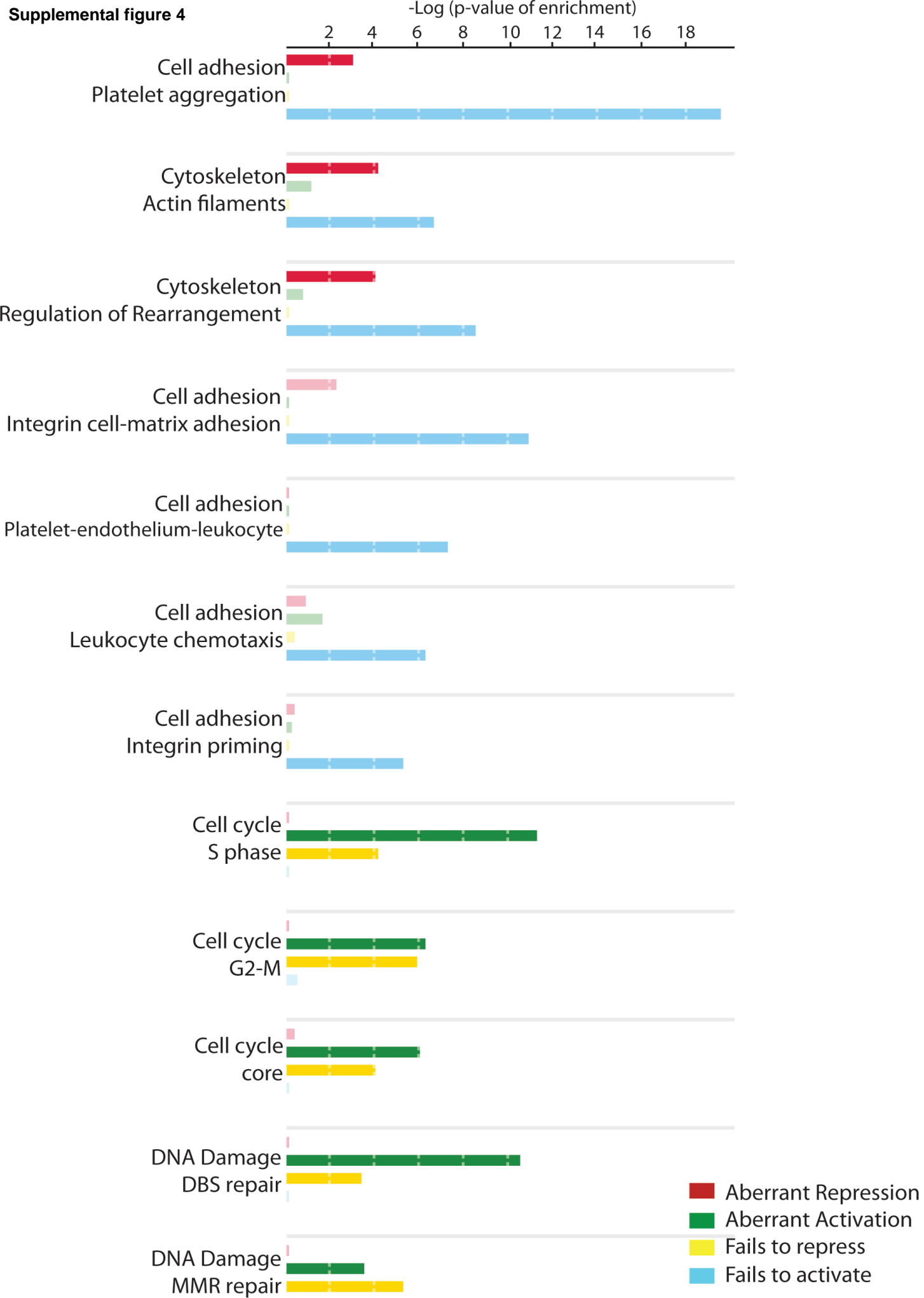
V205G

		Diff. expressed	
		No	Yes
Bound	No	22441	1101
	Yes	6702	610

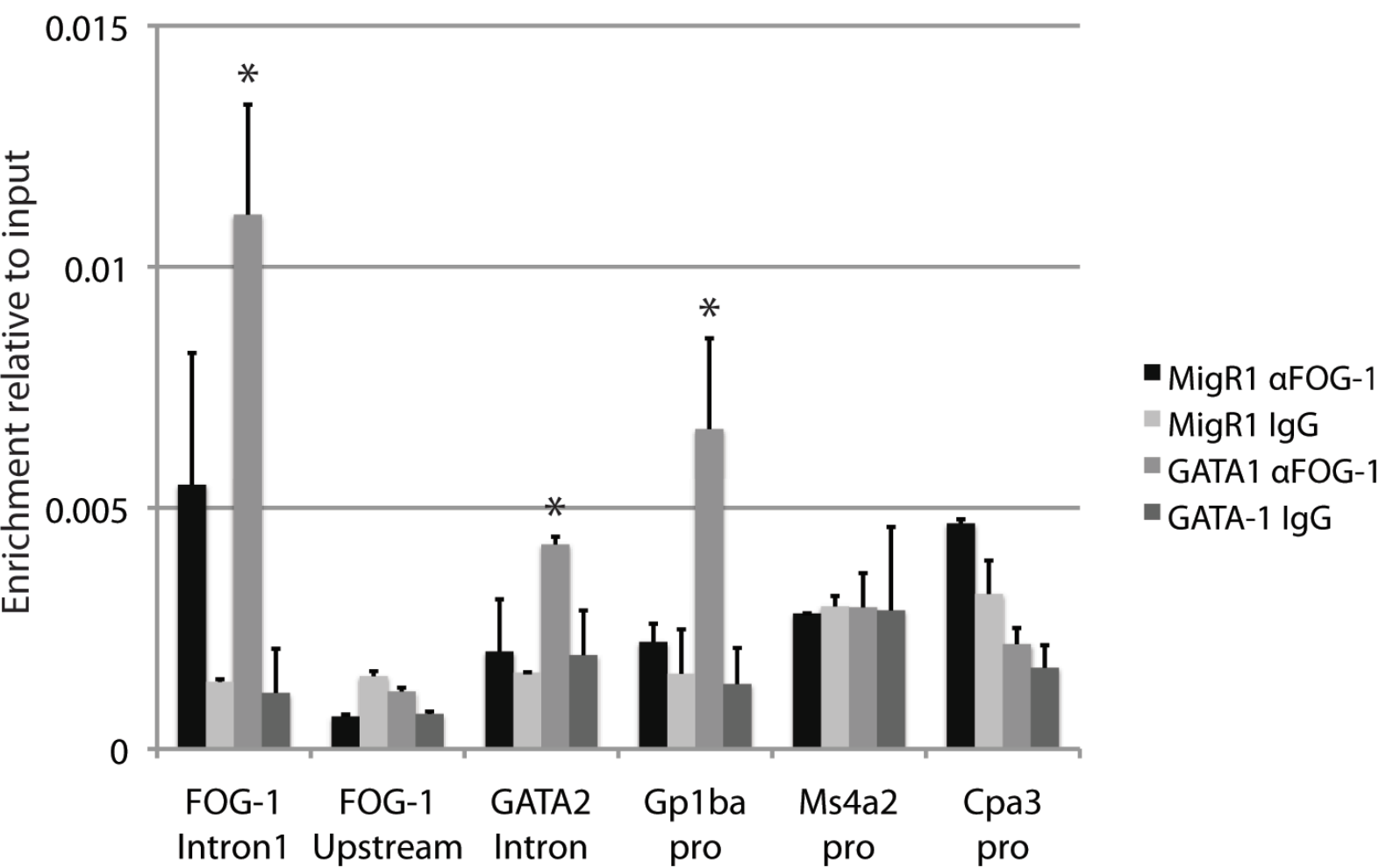
χ -squared= 142.40; p-value<2.2E-16



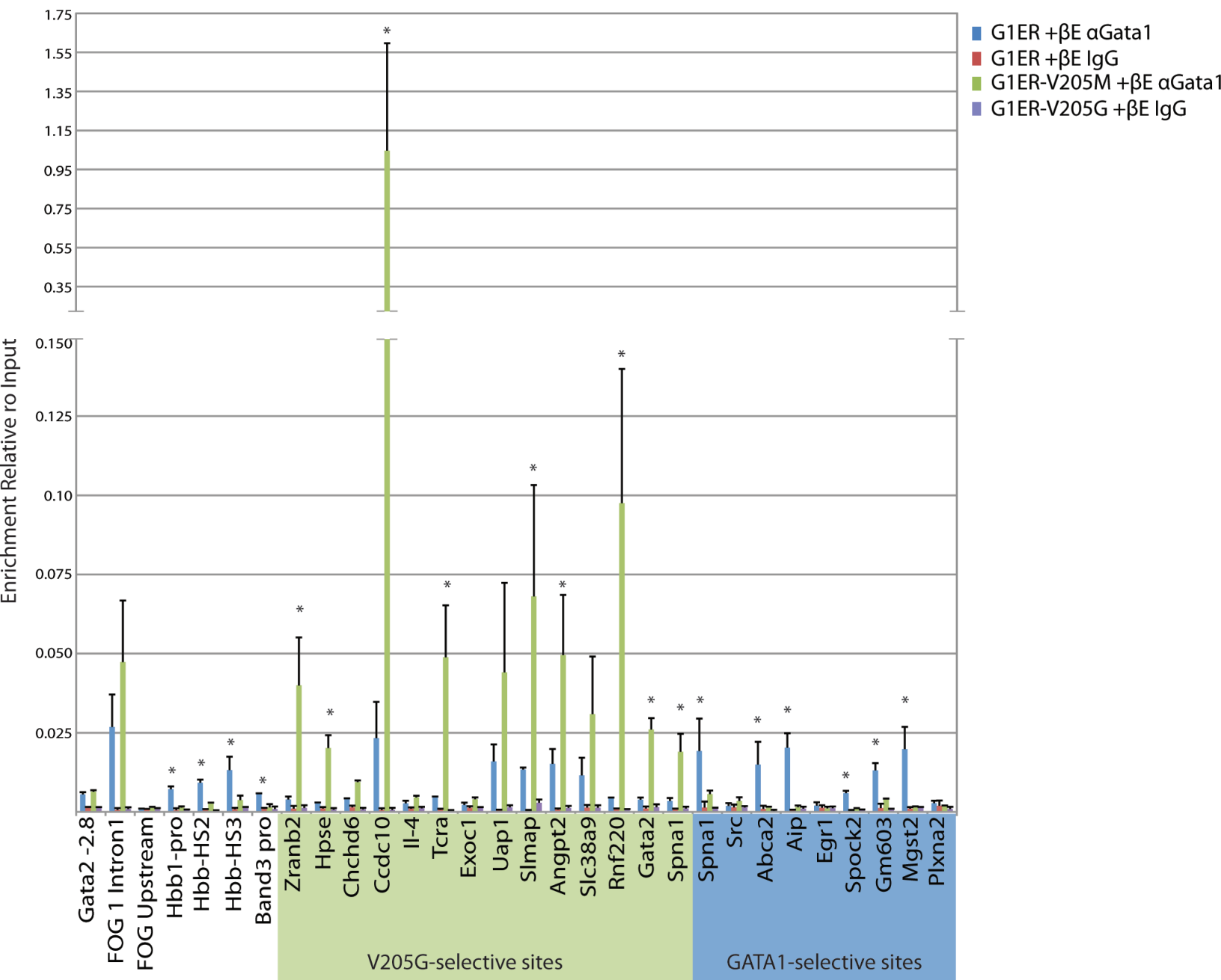
Supplemental figure 4



Supplemental figure 5



Supplemental figure 6



Chlon et al

Supplemental Figure legends and tables

Supplemental Figure 1 (related to main figure 2): ChIP-Seq for GATA-1 and GATA-

1V205G in G1ME cells. (A) GATA-1 and GATA-1V205G occupy control sites, including the *Itga2b* and *Gplba* promoters and the introns of *Zfpml*. (B) Each GATA-1 and GATA-1V205G peak was placed into a category based on its location relative to the nearest TSS: upstream (2-100kb), promoter (>2kb upstream), intragenic (exon or intron), downstream (<100kb), or gene desert (>100kb). Peaks within genes were further categorized for their presence in intron 1, 2, 3, 4, or ≥ 5 or an exon. Compared to a control set of peaks, GATA-1 and GATA-1V205G peaks are significantly enriched at promoters and intragenic regions and depleted at downstream and gene desert regions. (C) Each GATA1-selective and V205G-selective peak was placed into a category based on its location relative to the nearest TSS: upstream (2-100kb), promoter (<2kb upstream), intragenic (exon or intron), downstream (<100kb), or gene desert (>100kb). Peaks within genes were further categorized for their presence in intron 1, 2, 3, 4, or ≥ 5 or an exon. A control set of peaks was produced by randomly choosing a location on the same chromosome as every peak in the GATA-1 dataset. (D) A histogram of the frequency of GATA-1 and GATA-1V205G peaks at locations relative to the TSS. Bin size is 50bp.

Supplemental Figure 2 (related to main figure 3): *De Novo* motif analysis reveals deficiency of ETS motifs at the V205G-selective sites. (A) The first and second results from *de novo* motif analysis with DREME on 200bp of sequence surrounding the peak summits of the GATA-1 and GATA-1V205G peaks. (B) A chart of the first six motifs output from *de novo* motif analysis

with MEME on 200bp of sequence surrounding the peak summits of the GATA1-selective and V205G-selective peaks.

Supplemental Figure 3 (related to main figure 5): GATA-1 and GATA-1V205G

differentially expressed genes are enriched for genes that are bound. (A) Overlap of the genes that are differentially expressed by GATA-1 relative to MigR1-infected cells with genes that are bound by GATA-1 and V205G. Chi-square analysis indicates that the differentially expressed genes are enriched for GATA-1 bound genes. (B) A heatmap of the relative expression compared to MigR1 control of the genes that are bound and differentially expressed by GATA-1. The occupancy of each gene by GATA-1V205G is indicated in orange in an adjacent column. (C) Overlap of the genes that are differentially expressed by GATA-1V205G relative to MigR1-infected cells with genes that are bound by GATA-1 and V205G. Chi-square analysis indicates that the differentially expressed genes are enriched for GATA-1V205G bound genes. (D) A heatmap of the relative expression compared to MigR1 control of the genes that are bound and differentially expressed by V205G. The occupancy of each gene by GATA-1 is indicated in orange in an adjacent column.

Supplemental Figure 4 (related to main figure 4): GATA-1 requires interaction with FOG

to induce megakaryocyte differentiation pathways and inhibit cell cycle. Gene ontology analysis on the four groups of differentially expressed genes identified by hierarchical clustering. The negative log of the p-value of enrichment for each of the four groups is displayed for the 12 most enriched ontologies.

Supplemental Figure 5 (related to main figure 6): FOG-1 is not bound at V205G-selective sites. Chromatin immunoprecipitation for FOG-1 in MigR1 and GATA1-transduced G1ME cells. Means \pm SD are depicted. Asterisks mark sites where enrichment is significant compared to control IgG.

Supplemental Figure 6 (related to main figure 1): The GATA-FOG interaction regulates GATA-1 chromatin occupancy in erythroid cells. Chromatin immunoprecipitation for GATA-1 in G1ER cells expressing wild-type and V205M mutant GATA-1 treated with β -estradiol for 24h. The V205G-selective and GATA1-selective sites are highlighted in green and blue, respectively. Means \pm SD are depicted.

Supplementary Table 1. Statistics for ChIP-Seq datasets. Three biological replicate ChIPs were sequenced for each factor. Three total input samples from parental G1ME cells were sequenced.

Sample	Sequencing Reads		Peak Calls	Bound Genes
	Mappable	Unique		
Gata1	35,278,756	19,589,030	12,736	7,912
Gata1 ^{V205G}	37,398,428	19,383,347	11,696	7,312
INPUT	55,715,331	47,452,331	--	--

Supplementary Table 2. Overlap of the GATA-1 and GATA-1V205G ChIP-Seq peaks with H3K4me3 ChIP-Seq peaks from parental G1ME cells.

	H3K4me3 peaks	X ² Test against GATA-1 peaks
GATA-1	33% (4310)	
V205G	27% (3134)	
GATA1-selective	28% (402)	12.94 (p=0.0003)
V205G-selective	10% (152)	253.14 (p=5.37E-57)

Supplemental table 3

GeneID	chromosome	strand	gene start	gene end	region	distance
Sox17	chr1	-	4481008	4483816	Gene Desert	-108807
Sox17	chr1	-	4481008	4483816	Gene Desert	-129891
Mrpl15	chr1	-	4764014	4775768	Gene Desert	131569
Mrpl15	chr1	-	4764014	4775768	Downstream	29470
Tcea1	chr1	+	4847894	4887985	Promoter	-188
Tcea1	chr1	+	4848118	4880877	gene	16112
Rb1cc1	chr1	+	6204742	6265656	Promoter	-242
Sgk3	chr1	+	9788210	9892649	Upstream (10	-17619
Cops5	chr1	-	10014911	10027979	gene	63
Sulf1	chr1	+	12682609	12850453	Upstream (10	-47763
Prdm14	chr1	-	13103508	13115937	Upstream (10	-28843
Prdm14	chr1	-	13103508	13115937	Gene Desert	-102666
AK171953	chr1	-	13333668	13337135	Downstream	87305
AK171953	chr1	-	13333668	13337135	Downstream	79443
AK171953	chr1	-	13333668	13337135	Upstream (2-	-5978
Ncoa2	chr1	-	13129239	13362520	gene	5121
Tram1	chr1	-	13554782	13579945	gene	15540
Lactb2	chr1	-	13615979	13650590	gene	25
Eya1	chr1	-	14159038	14294651	gene	29803
Eya1	chr1	-	14159038	14294651	Upstream (10	-69069
Eya1	chr1	-	14159038	14294651	Gene Desert	-179962
Terf1	chr1	+	15795738	15833510	gene	511
Rdh10	chr1	+	16095962	16122631	Downstream	49151
Ube2w	chr1	-	16530870	16609367	gene	7205
Ly96	chr1	+	16678536	16699680	gene	10458
Jph1	chr1	-	16987544	17087879	gene	86994
AK205147	chr1	+	20793092	20793462	Upstream (2-	-7843
Efhc1	chr1	+	20941768	20980920	gene	6774
Efhc1	chr1	+	20941768	20980920	gene	21969
Efhc1	chr1	+	20941768	20980920	Downstream	42728
Tram2	chr1	-	20991458	21069306	gene	26792
Tram2	chr1	-	20991458	21069306	gene	13890
Tram2	chr1	-	20991458	21069306	Upstream (10	-10876
Tram2	chr1	-	20991458	21069306	Upstream (10	-22236
Gsta3	chr1	+	21230669	21255640	Downstream	26156
CS444445	chr1	+	23298552	23298575	Upstream (2-	-8955
Ogfrl1	chr1	-	23373262	23390014	Promoter	-448
Smap1	chr1	-	23851684	23916575	gene	9966
Smap1	chr1	-	23855633	23916575	Upstream (2-	-5974
Smap1	chr1	-	23852465	23929128	gene	483
Lmbrd1	chr1	+	24685764	24823146	Promoter	-163
4930521A18I	chr1	+	31279683	31281132	Gene Desert	155879
Prim2	chr1	-	33510655	33726603	gene	19797
Zfp451	chr1	-	33818597	33871272	gene	582
Bpag1	chr1	+	34344031	34365497	Upstream (10	-24975
Ccdc115	chr1	-	34493520	34496517	Downstream	4546
Imp4	chr1	+	34496744	34502592	Promoter	-69
AK016244	chr1	-	34534219	34542456	gene	980