Supplementary Fig. 1



Supplementary Figure 1. Pairwise differences between hESC lines in RLGS DNA fragments present in the lymphocyte "master" profile. Total differences and differences only in non-polymorphic (NP) loci are shown.

Supplementary Table1. Sequence characteristics of identified loci.

RLGS	Chromosome	Unigene ID/	Identity	CpG	Repeats	Gene	hESC	Differentiated	Tumour
fragment		GenBank		Island [#]	-	Ontology	expression	cell Expression	methylation
ĪD		Accession				Annotation	(references)	(references)	(references)
1C09	7q34	Hs.512833	<i>ZC3HAV1L</i> (zinc finger CCCH-type antiviral 1- like)	5'	None	Unknown	P ^(1, 2)	NL ⁽²⁾ , FF ⁽²⁾	M ^{(3)-G}
2B30	17q25.1	Hs.250723	<i>MXRA7</i> (matrix- remodelling associated 7, transmembrane anchor protein 1)	5'	Low complexity	Transport		EB ⁽⁴⁾ ,CD ⁽⁴⁾	
2C29	6q16.3	Hs.143410	CCNC (cyclin c)	None	SINE, LTR, Low complexity	Cell cycle	P ^(1, 2, 5-11)	NL ⁽²⁾ , FF ⁽²⁾	M ^{(12)-HNSCC}
2D10	5p15.33	AC122719	RP11-661C8 on chromosome 5	Present	None	Unknown			M ^{(13)-AML} ,M ^{(14)-MB}
2D24	8p21.3	Hs.106124	<i>EPB49</i> (erythrocyte membrane protein 4.9, dematin)	5'	SINE, MER	Protein binding			
2D25	10q26.3	Hs.501578	SPRN (shadow of prion protein)	5'	Low complexity	Unknown	P ^(1, 2)		
2D34	8q21.13	Hs.434957	<i>ZNF704</i> (zinc finger protein 704)	Intronic	Low complexity	Nucleic acid binding			M ^{(14)-MB}
2D40	6p21.1	Hs.311190	<i>MRP14</i> (mitochondrial ribosome protein L14)	5'	MER, Microsatellite	Protein biosynthesis	P ⁽¹⁾		
2D50	22q13.2	Hs.553531	<i>CGI-96</i> (gastric cancer antigen Zg14)	5'	LINE, SINE, LTR	Nucleic acid binding	P ^{(1, 4, 5)-} H9,(11)	EB ⁽⁴⁾ ,CD ⁽⁴⁾ FF ⁽²⁾	M ^{(3)-G}
2D57	22q11.1	Hs.474123	<i>CECR7</i> (cat eye syndrome critical region 7)	5'	LINE, SINE	Metabolism	A ⁽⁵⁾		
2D69	17q23.3	Hs.355264	<i>CYB561</i> (cytochrome b- 561 isoform 1)	5'	SINE, Microsatellite	Transport	P ^(1, 6)	K ⁽⁶⁾	M ^{(12, 15)-HNSCC}
2E03	3p21.33	Hs.19385	<i>ABHD5</i> (abhydrolase domain containing 5, CGI-58)	5'	Low complexity	Metabolism	P ^{(1)-BG02,(11)}	EB ⁽¹⁶⁾ ,K ⁽⁶⁾	
2E24	22q13.1	Hs.270285	SLC16A8 (solute carrier	5'	Low	Transport	P ^(4, 6, 7, 16, 17)		M ^{(13)-AML} , M

			16, MCT3)		complexity		A ⁽⁵⁾		^{(14)-MB} M ^{(3)-G} , M ^{(18)-LC} M ⁽¹⁹⁾⁻ CLL
2E61	17p11.2	Hs.190284	<i>RAI1</i> (retinoic acid induced 1)	None	LINE,SINE, Microsatellite	Regulation of transcription	P ^(1, 2, 8, 16, 20)	EB ⁽¹⁶⁾ ,B ⁽⁸⁾ FF ⁽²⁾ ,NL ⁽²⁾	M ^{(14)-MB} , M (18)-LC M ^{(19)-CLL}
2F57	2q11.1	Hs.469316	<i>LOC643085</i> (hypothetical protein)	5'	SINE,MER	Unknown	A ⁽⁵⁾		M ^{(19)-CLL}
3B24	2q14.2	AA016272	EST-unknown	None	Low complexity	Unknown			
3B36	2p22.2	Hs.154654	<i>CYP1B1</i> (cytochrome P450, family 1, subfamily B)	5'	Low complexity	Metabolism	P (1, 2, 4, 16)- H1/H7/H9,(6, 11, 20) A ^(5, 16) -HES2	EB ^(4, 11, 16, 20) ,B ⁽⁶⁾ , FF ⁽²⁾ ,NL ⁽²⁾	$\begin{array}{c} M^{(13)-AML}, M^{(14)-MB}, M^{(3)-G}, \\ M^{(18)-LC}, M^{(19)-}, \\ CLL, M^{(13)-}, \\ CC, G, AML, PE \end{array}$
3B46	2p13.2	BX108614	EST-unknown	None	SINE	Unknown			M ^{(15)-CC,G,AML}
3C01	18p11.21	Hs.136295	<i>GNAL</i> (GTP binding protein alpha-olf subunit)	5'	None	Signal transduction	P ^(16, 20) A ⁽⁵⁾	EB ⁽¹⁶⁾	M ^{(13)-AML} , M ^{(18)-LC} , M ⁽¹⁹⁾⁻ CLL
3C36	19p13.2	Hs.159360	<i>PTGER1</i> (prostaglandin E receptor 1,EP1)	5'	None	Signal transduction	A ⁽⁵⁾		
3C38	7q32.3	Hs.44693	MKLN1 (muskelin 1)	5'	Low complexity	Signal transduction	P (2, 16)- H1/H7/H9,(10)- BG02,(11) A ^(5, 16) -HES2	FF ⁽²⁾ ,NL ⁽²⁾	
3C40	7q36.3	Hs.490789	<i>PTPRN2</i> (protein tyrosine phosphatase receptor type N, polypeptide 2)	Intronic	None	Signal transduction	P (4, 16)- HES2,(8) A ^{(5, 16)-} H1/H7/H9	EB ⁽⁴⁾ ,B ⁽⁸⁾	M ^{(19)-CLL}
3C57	13q14.11	Hs.577960	FLJ42392 (hypothetical gene supported by AK124383)	Present	LINE	Unknown			M ^{(18)-AML} M ^{(19)-CLL}
3D10	1p36.11	Hs.181163	<i>HMGN2</i> (high mobility group nucleosomal binding domain 2)	5'	None	Nucleic acid binding	P ^(2, 9, 10, 21)	EB ⁽⁴⁾ ,CD ⁽⁴⁾ FF ⁽²⁾ ,NL ⁽²⁾	
3D23	10q11.22	Hs.538438	<i>LOC</i> 653468 (similar to	None	SINE,LINE,	Unknown	$P^{(5,(2, 6))}$	EB ⁽⁴⁾ ,CD ⁽⁴⁾	

			centaurin, gamma-like family, member 1)		Microsatellite			FF ⁽²⁾ ,NL ⁽²⁾	
3D24	17q23.2	Hs.121592	AP1S2 (adaptor-related protein complex1, sigma 2 subunit)	None	LINE,SINE, MER, Low complexity	Transport	P (1, 2, 5, 16),- H1/H7/H9,(6, 8),(20) A ⁽¹⁶⁾ -HES2	EB ⁽¹⁾ ,B ⁽⁶⁾ B-E-FF ⁽⁸⁾ FF ⁽²⁾ ,NL ⁽²⁾	M ^{(3)-G}
3E25	16q12.1	BG430697	EST-unknown	Present	SINE,LINE, LTR	Unknown			
3E68	12q13.13	Hs.533782	<i>KRT8</i> (keratin 8)	Present	SINE, Microsatellite	Protein binding	P ^{(1, 2, 5, 6, 8-10,} 16, 17, 20)	EB ^(4, 10, 11, 20) ,FF ⁽²⁾ ,NL ⁽²⁾	M ^{(19)-CLL}
3F50	10q24.31	Hs.89583	<i>TLX1</i> (T-cell leukemia homeobox 1)	Present	SINE	Transcription factor			M ^{(18)-LC} ,M ⁽¹³⁾⁻ AML M ⁽¹⁵⁾⁻ CC,AML,TC
3F64	4p16.3	Hs.369607	<i>GAK</i> (cyclin G associated kinase)	5'	SINE,LINE, MER,LTR, Microsatellite	Cell cycle	P ^(1, 2, 4-6, 16)	EB ⁽⁴⁾ , FF ⁽²⁾ NL ⁽²⁾	
4B08	12q23.2	Hs.382110	<i>LOC283432</i> (hypothetical protein)	5'	None	Unknown			
4C16	16q24.3	Hs.534773	<i>AFG3L1</i> (AFG3 ATPase family 3-like 1 isoform 1)	5'	SINE, Low complexity	ATP binding	P ⁽¹⁾		M ^{(13)-AML}
4C17	5q12.3	BQ771794	EST-unknown	Present	MER	Unknown			
4D08	17q25.3	Hs.440932	SEPT9 (MLL septin-like fusion)	None	LINE,SINE, LTR, Microsatellite	Cell cycle	P ^{(4),(10)-BG02}	EB ⁽⁴⁾	M ^{(3)-G} M ^(18)-L) , M ⁽²²⁾⁻ NS
4D12	11q13.1	Hs.502814	<i>LRP16</i> (low density lipoprotein receptor related protein 16)	None	LINE,SINE	Unknown	P ^{(1),(16)-} H1/H7/H9,(8) A ^{(16)-HES2}	EB ^{(16),(11)} ,B ⁽⁸⁾	M ^{(18)-AML} ,M (3)-G M ^{(19)-CLL} M ^{(15)-G,AML,TC} M ^{(22)-NS}
4E01	14q32.32	Hs.37712	<i>C14orf73</i> (chromosome 14 open reading frame 73)	5'	SINE,MER, Microsatelite	Unknown			M ^{(12)-HNSCC} M ^{(15)-TC}
4E16	4q34.1	Hs.127407	<i>GALNT7</i> (N'acetylgalactosaminyl- transferase 7)	5'	LINE,SINE, Low complexity	Metabolism	P ^{(1, 2, 4, 5),(16)-} H1/H7/H9,(6, 8),(10) -	EB ⁽⁴⁾ ,CD ⁽⁴⁾ FF ^(2, 8) , NL ⁽²⁾	M ^{(18)-BC} , M ^{(19)-CLL}

							H1/H7/H9,22)		
							A (16)-HES2		
4F32	10q26.3	Hs.591374	<i>EBF3</i> (early b-cell factor	5'	Low	Regulation			
	_		3)		complexity,	of			
					Microsatellite	transcription			
4F58	12q21.1	Hs.27214	KCNC2 (shaw-related	5'	LTR,	Transport	A ⁽⁵⁾		M ^{(18)BC,G,AML}
			voltage-gated potassium		Microsatellite				M ^{(19)-CLL}
			channel 2)		Low				
					Complexity				
5B03	3q29	Hs.274539	BDH (3'hydroxybutyrate	5'	SINE,	Metabolism	$P^{(6, 8), (20), (11)}$	$B^{(8),(20)}$	
			dehydrogenase)		Microsatellite		A ⁽⁵⁾		
5E27	7p15.3	Hs.432390	DNAH11 (dynein, heavy	5'	MER,	ATP binding	$P^{(8)}_{(5)}$		
			polypeptide 11)		Microsatellite		A ⁽⁵⁾		
2B61 [*]		U13369	Ribosomal RNA repeat		Tandem				
			(28S and 3' external						
*			transcribed spacer)						
$2G88^{*}$		U13369	18 S ribosomal RNA		Tandem				
			repeat	_					
3E64		U13369	Ribosomal RNA repeat		Tandem				
			(3' external transcribed						
2025*		11100.00	spacer)						
3G35		U13369	28 S ribosomal RNA		Tandem				
4CD*		1112260	repeat	_					
4CB		U13369	28 S ribosomal RNA		Tandem				
4020*		1112260	repeat						
4D28		013369	28 S FIDOSOMAI RINA		Tandem				
4E22*		U12260	28 S ribosomal DNA		Tandam				
4E32		015509	28 S HOOSOIIIai KINA		Tanuem				
4F33*		U13369	28 S ribosomal RNA		Tandem				
HL33		015507	repeat		Tanacin				
4F42*		U13369	28 S ribosomal RNA		Tandem				
1212		010007	repeat		Tundonn				
4E43*		U13369	28 S ribosomal RNA		Tandem				
_			repeat		-				
5B25 [*]		U13369	Ribosomal RNA repeat		Tandem				

		(intergenic spacer)			
5E55*	U13369	Ribosomal RNA repeat (3' External transcribed spacer and intergenic spacer)	Tandem		
5G01*	U13369	Ribosomal RNA repeat (internal transcribed spacer 2)	Tandem		
6G12 [*]	U13369	Ribosomal RNA repeat (intergenic spacer)	Tandem		

*= ribosomal DNA repeats identified according to Kuick and colleagues (23); $^{\#}CpG$ islands were classified as 5' if spanning the region upstream of the transcription start site or first exon, as *Intronic* if spanning internal introns and finally as *Present* if location could not be identified within a gene or EST; P= present ; A= absent; EB= embryoid body; CD= chemical differentiation (retinoic acid/ mitogen or DMSO); K= keratinocytes; B=differentiated blood; E= endothelial cells; FF= fetal fibroblasts; NL= non lineage differentiated cells; M= methylated; D= demethylated; M or D= methylated or demethylated; HNSCC= head and neck squamous cell carcinoma; AML= acute myeloid leukemia MB= medulloblastoma; G= glyoma; LC=lung cancer; CLL= chronic lymphocytic leukaemia; CC= colon cancer; PE= primitive ectodermal tumours; TC= testicular carcinomas; NS= nonseminomas; BC= breast cancer.

Supplementary Table 2. Unidentified loci present in our lymphocyte "master profile" that vary between or within hESC lines and have previously been identified as aberrantly methylated in tumours.

RLGS	Methylated in tumours
ID	(references)
2B32	HNSCC ⁽¹²⁾
2D63	AML ⁽¹³⁾
2D64	HNSCC ⁽¹²⁾ , MB ⁽¹⁴⁾ , CC, HNSCC, AML, PE, TC ⁽¹⁵⁾
2F46	MB ⁽¹⁴⁾ , CC, G, AML, PE, TC ⁽¹⁵⁾
2F56	MB ⁽¹⁴⁾ , BC, G, AML, PE ⁽¹⁵⁾
3B01	HNSCC ⁽¹²⁾
3E51	HNSCC ⁽¹²⁾
3F61	HNSCC ⁽¹²⁾
4B07	AML ⁽¹²⁾
4C08	HNSCC ⁽¹²⁾

HNSCC= head and neck squamous cell carcinoma; AML= acute myeloid leukaemia; MB= medulloblastoma; G= glyoma; CC= colon cancer; PE= primitive ectodermal tumours; TC= testicular carcinomas; BC= breast cancer; EC= teratocarcinoma.

References

- 1. Brandenberger, R., Khrebtukova, I., Thies, R.S., Miura, T., Jingli, C., Puri, R., Vasicek, T., Lebkowski, J. and Rao, M. (2004) MPSS profiling of human embryonic stem cells. *BMC Dev Biol*, 4, 10.
- Skottman, H., Mikkola, M., Lundin, K., Olsson, C., Stromberg, A.M., Tuuri, T., Otonkoski, T., Hovatta, O. and Lahesmaa, R. (2005) Gene expression signatures of seven individual human embryonic stem cell lines. *Stem Cells*, 23, 1343-56.
- 3. Zardo, G., Tiirikainen, M.I., Hong, C., Misra, A., Feuerstein, B.G., Volik, S., Collins, C.C., Lamborn, K.R., Bollen, A., Pinkel, D. *et al.* (2002) Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors. *Nat Genet*, 32, 453-8.
- 4. Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J., Li, Y., Xu, C., Fang, R., Guegler, K. *et al.* (2004) Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. *Nat Biotechnol*, 22, 707-16.
- 5. Abeyta, M.J., Clark, A.T., Rodriguez, R.T., Bodnar, M.S., Pera, R.A. and Firpo, M.T. (2004) Unique gene expression signatures of independentlyderived human embryonic stem cell lines. *Hum Mol Genet*, 13, 601-8.
- Golan-Mashiach, M., Dazard, J.E., Gerecht-Nir, S., Amariglio, N., Fisher, T., Jacob-Hirsch, J., Bielorai, B., Osenberg, S., Barad, O., Getz, G. *et al.* (2005) Design principle of gene expression used by human stem cells: implication for pluripotency. *Faseb J*, 19, 147-9.
- 7. Sato, N., Sanjuan, I.M., Heke, M., Uchida, M., Naef, F. and Brivanlou, A.H. (2003) Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev Biol*, 260, 404-13.
- 8. Sperger, J.M., Chen, X., Draper, J.S., Antosiewicz, J.E., Chon, C.H., Jones, S.B., Brooks, J.D., Andrews, P.W., Brown, P.O. and Thomson, J.A. (2003) Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc Natl Acad Sci U S A*, 100, 13350-5.
- 9. Zeng, X., Miura, T., Luo, Y., Bhattacharya, B., Condie, B., Chen, J., Ginis, I., Lyons, I., Mejido, J., Puri, R.K. *et al.* (2004) Properties of pluripotent human embryonic stem cells BG01 and BG02. *Stem Cells*, 22, 292-312.
- 10. Bhattacharya, B., Cai, J., Luo, Y., Miura, T., Mejido, J., Brimble, S.N., Zeng, X., Schulz, T.C., Rao, M.S. and Puri, R.K. (2005) Comparison of the gene expression profile of undifferentiated human embryonic stem cell lines and differentiating embryoid bodies. *BMC Dev Biol*, 5, 22.
- 11. Enver, T., Soneji, S., Joshi, C., Brown, J., Iborra, F., Orntoft, T., Thykjaer, T., Maltby, E., Smith, K., Dawud, R.A. *et al.* (2005) Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells. *Hum Mol Genet*, 14, 3129-40.
- 12. Smiraglia, D.J., Smith, L.T., Lang, J.C., Rush, L.J., Dai, Z., Schuller, D.E. and Plass, C. (2003) Differential targets of CpG island hypermethylation in primary and metastatic head and neck squamous cell carcinoma (HNSCC). *J Med Genet*, 40, 25-33.
- 13. Rush, L.J., Dai, Z., Smiraglia, D.J., Gao, X., Wright, F.A., Fruhwald, M., Costello, J.F., Held, W.A., Yu, L., Krahe, R. *et al.* (2001) Novel methylation

targets in de novo acute myeloid leukemia with prevalence of chromosome 11 loci. *Blood*, 97, 3226-33.

- 14. Fruhwald, M.C., O'Dorisio, M.S., Dai, Z., Tanner, S.M., Balster, D.A., Gao, X., Wright, F.A. and Plass, C. (2001) Aberrant promoter methylation of previously unidentified target genes is a common abnormality in medulloblastomas--implications for tumor biology and potential clinical utility. *Oncogene*, 20, 5033-42.
- 15. Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomaki, P., Lang, J.C. *et al.* (2000) Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet*, 24, 132-8.
- 16. Wei, C.L., Miura, T., Robson, P., Lim, S.K., Xu, X.Q., Lee, M.Y., Gupta, S., Stanton, L., Luo, Y., Schmitt, J. *et al.* (2005) Transcriptome profiling of human and murine ESCs identifies divergent paths required to maintain the stem cell state. *Stem Cells*, 23, 166-85.
- Bhattacharya, B., Miura, T., Brandenberger, R., Mejido, J., Luo, Y., Yang, A.X., Joshi, B.H., Ginis, I., Thies, R.S., Amit, M. *et al.* (2004) Gene expression in human embryonic stem cell lines: unique molecular signature. *Blood*, 103, 2956-64.
- Dai, Z., Lakshmanan, R.R., Zhu, W.G., Smiraglia, D.J., Rush, L.J., Fruhwald, M.C., Brena, R.M., Li, B., Wright, F.A., Ross, P. *et al.* (2001) Global methylation profiling of lung cancer identifies novel methylated genes. *Neoplasia*, 3, 314-23.
- 19. Rush, L.J., Raval, A., Funchain, P., Johnson, A.J., Smith, L., Lucas, D.M., Bembea, M., Liu, T.H., Heerema, N.A., Rassenti, L. *et al.* (2004) Epigenetic profiling in chronic lymphocytic leukemia reveals novel methylation targets. *Cancer Res*, 64, 2424-33.
- 20. Cai, J., Chen, J., Liu, Y., Miura, T., Luo, Y., Loring, J.F., Freed, W.J., Rao, M.S. and Zeng, X. (2006) Assessing self-renewal and differentiation in human embryonic stem cell lines. *Stem Cells*, 24, 516-30.
- 21. Richards, M., Tan, S.P., Tan, J.H., Chan, W.K. and Bongso, A. (2004) The transcriptome profile of human embryonic stem cells as defined by SAGE. *Stem Cells*, 22, 51-64.
- 22. Smiraglia, D.J. and Plass, C. (2002) The study of aberrant methylation in cancer via restriction landmark genomic scanning. *Oncogene*, 21, 5414-26.
- 23. Kuick, R., Asakawa, J., Neel, J.V., Kodaira, M., Satoh, C., Thoraval, D., Gonzalez, I.L. and Hanash, S.M. (1996) Studies of the inheritance of human ribosomal DNA variants detected in two-dimensional separations of genomic restriction fragments. *Genetics*, 144, 307-16.