QTL analyses of lineage negative mouse bone marrow cells labelled with Sca-1 and c-kit

Mays Jawad, Clare Cole, Abigail Zanker, George Giotopoulos, Simon Fitch, Christopher J. Talbot and Mark Plumb

Department of Genetics, University of Leicester, Leicester LE1 7RH, U.K.

Correspondence: Chris Talbot
Department of Genetics, University of Leicester,
University Road, Leicester LE1 7RH, U.K.
Tel: +44 (0)116 252 3433
Fax: +44 (0)116 252 3378
E.mail: cjt14@le.ac.uk

Running title: QTL for stem cell frequency

Keywords: Stem cell, quantitative trait loci, flow cytometry

Abstract

Differences in the number of functionally and/or phenotypically defined bone marrow cells in inbred mouse strains have been exploited to map quantitative trait loci (QTL) that determine the variation in cell frequency. To extend this approach to the differences in the stem/progenitor cell compartment in CBA/H and C57BL/6 mice, we have exploited the resolution of flow cytometry and the power of QTL analyses in 124 F2 mice to analyse lineage negative (Lin⁻) bone marrow cells according to the intensity of labelling with Sca-1 and c-Kit. In the Lin⁻ Sca-1⁺ c-Kit⁺ enriched population six quantitative trait loci (QTL) were identified – one significant and five suggestive. Whereas previous in vitro clonogenic, LTIC-IC, CAFC day 35, and flow cytometry each identified different QTL, our approach identified the same or very similar QTL at all three loci (Chromosome 1, 17 and 18) as well as QTL on chromosomes 6 and 10. *In silico* analyses implicate haematopoietic stem cell homing involving Cxcr4 and Cxcl12 as being the determining pathway. The mapping of the same or very similar QTLs in independent studies using different assay(s) suggests a common genetic determinant, and thus reinforces the biological and genetic significance of the QTL. These data also suggest that mouse bone marrow cell sub-populations can be functionally, phenotypically and genetically defined.

Introduction

The regulation of haemopoiesis relies on the proliferation and/or differentiation of bone marrow stem and progenitor cells which have been functionally defined using either *in vitro* clonogenic assays or *in vivo* repopulation assays in engrafted mice. Significant differences in normal haematopoiesis have been reported between inbred mouse strains, and these differences have enabled genetic studies to identify quantitative trait loci (QTL) that regulate these differences. For example, QTL that determine stem cell numbers (Henckaerts et al. 2002, 2004; De Haan and Van Zant 1997, 2000, 2002; Muller-Sieburg and Riblet 1996; Morrison et al. 1997, 2002), the frequency of progenitor cells and the relative proportion of mature blood red, T and B cells (Chen and Harrison 2002; Morrison et al. 2002; van Os et al. 2006; Peters et al. 2006; Valdar et al. 2006; Jawad et al. 2007a) have been identified. In a landmark study the gene underlying one of these QTL has recently been identified as *latexin* (Liang et al. 2007). Despite the hierarchical nature of haemopoiesis, there is little overlap in QTL that determine the frequencies of stem, progenitor and mature haemopoietic cells, suggesting that functionally defined mouse bone marrow cells can also be genetically defined.

In vitro bone marrow stem and immature progenitor cell assays are based on stroma-dependent long-term bone marrow cultures. The long-term culture initiating cell (LTC-IC) and Cobblestone Area Forming Cells (day 35 CAFC) have been shown to be related to the bone marrow repopulating stem cell in BXD RI mice (Ploemacher et al. 1991; Muller-Sieburg and Riblet, 1996; Sutherland et al. 1990; Petzer et al. 1996; Breems et al. 1994). As described in Table 1, the initial study using LTC-IC identified QTL on chromosome 1 (Muller-Sieburg and Riblet, 1996) and although this was missed in an initial CAFC Day 35 study (QTL on chromosome 18 only; de Haan and Van Zant, 1997), a more recent CAFC Day 35 study identified both QTL on chromosomes 1 and 18 (Geiger et al. 2001).

Stem and progenitor cells have also been phenotypically defined according to the expression of specific cell surface antigens such as Sca-1 (Ly6A/E), Thy-1, c-Kit, CD34 and Flk-2 (Christensen and Weissman, 2001; Morrison et al. 1995, 1997; Okada et al. 1992; Spangrude and Brooks, 1992). They do not express lineage

markers found in maturing blood cells (Lin⁻), are Thy-1.1^{lo} and are positive for the expression of Sca-1 and c-Kit. For example, Lin⁻ Thy-1.1^{lo} Sca-1⁺ or Lin⁻ CD34^{lo/-} Sca-1⁺ c-Kit⁺ bone marrow cells provides stable, long-term multi-lineage haemopoiesis in recipient mice (Uchida and Weissman, 1992; Osawa et al. 1996). The QTL that determine the frequencies of LTC-IC and day 35 CAFC are also unrelated to the (C57BL/Ka-Thy-1.1 X AKR/J)F2 chromosome 17 QTL that determine the frequency of Lin⁻ Thy-1.1^{low} Sca-1⁺ c-Kit⁺ cells analysed directly by flow cytometry (Morrison et al. 2002; Table 1). This suggests that the assays may be examining different cells within the stem/progenitor cell compartment, or in the case of *in vitro* long term assays, differences in their ability to produce colonies *in vitro*.

Although the evidence suggests that functionally defined bone marrow stem and progenitor cells can also be both phenotypically and genetically defined, the genetic relationship between apparently similar phenotypes is confused. Theoretically some if not all of the cells that give rise to LTC-IC and/or day 35 CAFC colonies should be positive for Sca-1 and c-Kit. Similarly, functionally defined LTC-IC and day 35 CAFC should be genetically related to Lin⁻ Thy-1.1^{low} Sca-1⁺ c-Kit⁺ cells. We have therefore analysed total Lin⁻ Sca-1⁺ c-Kit⁺ bone marrow cells in (CBA/H x C57BL/6)F2 mice directly by Fluorescence Activated Cell Scanning (FACS) and mapped QTL, three of which map to the same or similar chromosomal intervals that determine the frequency of LTC-IC (chromosome 1), day 35 CAFC (chromosome 18) and Lin⁻, Thy-1.1^{low}, Sca-1⁺, c-Kit⁺ (chromosome 17) cells (Muller-Sieburg and Riblet, 1996; de Haan and Van Zant, 1997; Morrison et al. 2002). This direct analysis by flow cytometry together with a genome-wide screen in F2 mice provides supportive evidence that functionally defined bone marrow cell sub-populations are also both phenotypically and genetically defined.

Materials and Methods

Mice

CBA/H mice were from the Harwell colony, and C57BL/6, DBA/2 mice were from Harlan UK Limited. (CBA/H x C57BL/6)F1 and F2 mice were bred in Leicester. The

animal studies were carried out under guidance issued by the MRC in 'Responsibility in the use of animals for medical research' (July 1993) and Home Office Project licences No PPL 80\1565 and 80\1564.

Flow cytometry analysis and Isolation of Primitive Haematopoietic progenitors

Bone marrow was flushed from the femurs of 8-12 week old mice and depleted of lineage positive (Lin⁺) cells using streptavidin-conjugated magnetic beads (QIAGEN, Northampton, U.K) and a panel of biotinylated anti CD3e, CD11b, CD45R, Ly-6G and Ly-6C, and TER-119 antibodies. Lin⁻ bone marrow cells were counted (Scharfe System Casey 1 machine) and then labelled with R-phycoerythrin- (R-PE) conjugated anti mouse *c-Kit* and fluorescein isothiocyanate-(FITC) conjugated anti Sca-1 or appropriate isotype controls. All antibodies were from Becton Dickinson-Pharmingen (Oxford, U.K.). The whole cell population acquisition and analysis was performed on a two-laser FACScalibur (Becton Dickinson) and data analysed using Cellquest software. Individual mice were analysed.

Background, autofluorescence and non-specific antibody labelling was determined using unlabelled bone marrow cells, or bone marrow cells labelled with appropriate rat IgG2 isotype controls, or single labelling with specific FITC Sca-1 or R-PE *c-Kit* antibodies.

Genetic analysis

Lin⁻ bone marrow cells from 124 F2 mice were analysed for Sca-1 and *c-Kit* expression by flow cytometry. If necessary, quantitative data from the flow cytometer were normalised (by square root transformation) for QTL analysis. DNA prepared from the tails of the F2 mice was genotyped using 126 polymorphic microsatellite markers as previously described (Boulton et al. 2003) in a genome-wide screen at ~20 cM intervals. The genotype and phenotype data for each of the 124 F2 mice were analysed by composite interval mapping using QTL Cartographer (Wang et al. 2001-2004). Significance linkage levels were calculated by permuting the data 10,000 times

and interpreted as determined by Lander and Kruglyak (1995). The calculated thresholds showed little variation between phenotypes, for LSK they were: Suggestive linkage LOD >2.5, significant linkage LOD >3.5, and highly significant linkage LOD > 5.8. 95% confidence intervals are based on one LOD support intervals (Lynch and Walsh 1998). QTL positions and effect sizes were estimated using multiple interval mapping. Megabase positions are based on genome build NCBI m36.

The three quantitative traits for which QTL were found are highly correlated (R1 vs Lin⁻, Sca-1⁺⁺ r=0.76; R1 vs Lin⁻, Sca-1⁻, c-Kit⁻ r=-0.65; Lin⁻, Sca-1⁺⁺ vs Lin⁻, Sca-1⁻, c-Kit⁻ r=-0.78) as expected, so no adjustment of significance thresholds was used.

In silico analysis

Candidate genes in the QTL were assessed against four criteria. Firstly, we interrogated the GeneNetwork database (www.genenetwork.org) for genes showing cis-regulation of expression in the QTL (Liang et al. 2007, Bystrykh et al. 2005), using data from the GNF haematopoietic cells on Affymetrix microarray U74Av2 for the BXD cross. Secondly, we searched the GNF SymAtlas (<http://symatlas.gnf.org/SymAtlas/>) for genes showing bone marrow expression greater than 3-fold the tissue median, which again uses data from Affymetrix U74 arrays (Su et al. 2004). Thirdly we searched the Mouse Genome Database for genes with mutants known to have haematopoietic phenotypes (Eppig *et al*, 2005). Fourthly the strain distribution patterns of SNPs across the loci were compared for the strains in the cross (data from Scripps Florida Mouse Database and WTCHG (Cervino *et al*, 2006; Valdar *et al*, 2006)) to identify chromosomal regions of divergent ancestry.

Results

Parental CBA/H and C57BL/6 mice

To enumerate the frequency of Lin⁻ cells according to the intensity of Sca-1 and/or c-Kit labelling in the parental CBA/H and C57BL/6 inbred mouse strains, Lin⁻ bone marrow cells were isolated by immunomagnetic depletion of the Lin⁺ cells. There was no significant difference in the recovery of total non-erythroid bone marrow cells (WBMC) or Lin⁻ cells (~3% of WBMC) from the two inbred mouse strains or the (CBA/H x C57BL/6)F1 hybrid (Jawad et al, 2007a). The Lin⁻ cells were labelled with FITC anti-Sca-1 and PE anti-c-Kit antibodies and all cells visualised by flow cytometry.

Gating strategy was selected such that >98% of unlabelled Lin⁻ bone marrow cells appeared in the 30 x 30 lower left cytogram quadrant and therefore positive fluorescent events were classified as events in the >30 quadrant limit (Figure 1A). Similar results were obtained using FITC and R-PE conjugated isotype control antibodies (data not shown). The labelling of Lin⁻ bone marrow cells with single R-PE c-Kit (Figure 1B) or FITC Sca-1 antibodies (Figure 1C) illustrate both the specificity of antibody binding and reinforce the 0-30 quadrant definition of unlabelled cells. As shown in Figures 1A-C, there is some background and autofluorescence, but altogether, using unlabelled cells, isotype controls and single specific antibody labelling of Lin⁻ cells, background fluorescence outside the defined 0-30 limits on the x and y axis was estimated as <3%.

Lin⁻ cells prepared from CBA/H (Figure 2A), C57BL/6 (Figure 2B) and DBA/2 (Figure 2C) bone marrow labelled with both c-Kit and Sca-1 antibodies further illustrate the specificity of the flow cytometric analyses. Compared to C57BL/6, CBA/H and DBA/2 Lin⁻ cells are enriched for a discrete subpopulation of Lin⁻Sca-1⁺c-Kit⁺ cells (L⁻S⁺K⁺; Figure 2A and C, highlighted) and this is well above the background. For convenience, cells that fluoresce with an intensity of between 30 and 10² are arbitrarily defined as (+), whereas those that fluoresce with an intensity of

$>10^2$ are defined as (++). Thus, the cytograms can be quantitatively analysed according to the nine areas defined by the relative fluorescence intensity (-, + or ++, Figure 2D) of the Sca-1 and c-Kit coordinates in addition to the gated (R1) discrete population of $\text{Lin}^- \text{Sca-1}^+ \text{c-Kit}^+$ cells (R1; Figure 2D).

As shown in Table 2, there is a statistically significant genotype-dependent difference in the frequency of the discrete subpopulation of gated $\text{L}^- \text{S}^+ \text{K}^+$ cells ($p = 0.0054$). It is worthy of note that the frequency of all Sca-1 positive cells that fluoresce intensely ($y > 10^2$; highlighted Figure 2B) is higher in C57BL/6 and DBA/2 than CBA/H. This is consistent with C57BL/6 mice that carry the Ly6^b haplotype giving higher Sca-1 expression than the Ly6^a haplotype of the CBA/H mice (Spangrude and Brooks, 1993). Also, the frequency of Sca-1^+ cells in F1 hybrid mice are similar to the C57BL/6 mice, suggesting that the higher intensity of the Sca-1^+ cells is determined by a dominant C57BL/6 genetic train (data not shown). Despite the Ly-6 haplotypes, CBA/H (and DBA/2) mice have a ~ 4 fold higher frequency of the discrete $\text{L}^- \text{S}^+ \text{K}^+$ cells than C57BL/6 mice. CBA/H mice are closely related to DBA/2 mice (Festing 1996), and the four-fold $\text{L}^- \text{S}^+ \text{K}^+$ cell frequency difference is the same order of magnitude as the higher frequency of stem cells in DBA/2 and CBA mice compared to C57BL/6 mice estimated using *in vitro* LTC-IC (Muller-Sieberg and Riblet 1996) and/or day 35 CAFC (de Haan and Van Zant 1997) *in vitro* clonogenic assays.

Despite the observed difference in the intensity/number of $\text{Sca-1}^{+/++}$ cells between the two parental strains, there is less of a difference in the total number of $\text{Lin}^- \text{Sca-1}^{++} \text{c-Kit}^{++}$ (Upper right quadrant in Figure 2D) cells between the two inbred strains, the difference is not statistically significant and both are comparable to that observed in the F1 hybrid (data not shown). This is due in part to the observed higher frequency of the discrete R1 $\text{L}^- \text{S}^+ \text{K}^+$ gated population (Table 2) in CBA/H mice than C57BL/6 mice, indicating that reduced difference in the total number of $\text{Lin}^- \text{Sca-1}^{++} \text{c-Kit}^{++}$ cells is a balance between higher Sca-1 labelling in the C57BL/6 mice and a higher frequency of the gated R1 ($\text{L}^- \text{S}^+ \text{K}^+$) cells in CBA/H mice.

Genetic analysis

Lin⁻ bone marrow cells from each of 124 (CBA/H X C57BL/6)F1 intercross F2 mice were analysed for Sca-1 and c-Kit labelling by flow cytometry, and genotyped in a genome wide screen at ~20 cM intervals with 126 polymorphic microsatellite markers. Quantitative data for each mouse and FACS area (Figure 2D) were analysed using the QTL Cartographer programme (Wang et al. 2001-2004). Areas of the flow cytogram that revealed genetic linkage above the threshold necessary for suggestive linkage, as calculated by permuting the data (LOD= 2.5), are summarised in Table 3 (Lander and Kruglyak 1995). As the quantitative traits used are correlated no correction was applied for multiple testing.

Composite interval mapping found a significant QTL on chromosome 17 that determines the frequency of double negative L⁻S⁻K⁻ cells, which maps to the HLA locus (Figure 5 and Table 3). Analysis of the same L⁻S⁻K⁻ cells identified a suggestive QTL on chromosome 19. No QTL were revealed for Lin⁻ c-Kit⁺ cells ($x > 30$), but a significant QTL on chromosome 15 was revealed for the intensely labelling Lin⁻ Sca-1⁺ cells (Lin⁻Sca-1⁺⁺; $y > 10^2$; Figures 2 & 4 and Table 3). The mouse Sca-1 gene maps to chromosome 15 so this result is consistent with the known Ly-6 haplotypes of the parental mouse strains (Spangrude and Brooks 1993).

Five QTL were revealed for the discrete population of L⁻S⁺K⁺ cells, four of which achieve the criteria of suggestive (LOD > 2.5; chromosomes 6 and 17, 18), and one is significant (LOD 3.5, chromosome 10) (Table 3 and Figure 3). Optimisation of final QTL models by Multiple Interval Mapping reveals an additional QTL on chromosome 1 at 62cM (an epistatic effect with the chromosome 17 QTL) and estimates the effect sizes of the QTL as being: chromosome 1 4.8%, chromosome 6 5.0% and 8.1%, chromosome 10 10.8%, chromosome 17 5.1%, chromosome 18 6.7%.

Two of the QTL map close to QTL that determine the frequency of stem cells in (BXD)RI mice using the LTC-IC (*Scf*², chromosome 1; ~53 cM)(Muller-Sieburg and Riblet 1996) and day 35 CAFC (chromosome 18; 19 cM) (de Haan and Van Zant, 1997) in vitro clonogenic stem cell assays (Table 1). The chromosome 17 QTL (HLA complex) was also revealed by flow cytometry of Lin⁻ Thy-1^{low} Sca-1⁺ c-Kit⁺ (AKR/J

x C57Bl/Ka)F2 bone marrow cells (Morrison et al. 2002). Interestingly, the chromosome 15 (Ly6) QTL was not revealed in the analysis of the discrete population of Lin⁻ Sca-1⁺ c-Kit⁺ cells (Table 3).

There is no evidence of a sex effect in these data; none of the traits are significantly different between males and females and introducing sex as a co-variate in the QTL analysis makes little difference (data not shown).

***In silico* analysis**

Each QTL contains many genes but *in silico* analyses were carried out to compile a list of candidate quantitative trait genes (QTG). To assess the specificity of our criteria we scored all protein-coding genes on mouse chromosome 2, for which no QTL were detected. Of the 1913 protein-coding genes 2.0% show *cis* regulation of transcription, 4.8% have bone marrow expression >3-fold above median and 4.1% have mutants with known haematopoietic phenotypes. Only one chromosome 2 gene (Cd44) meets all three of these criteria, and it also meets the fourth criterion, by being in a genomic region with SNPs informative for this cross. Thus although these criteria will miss candidate QTG e.g. those not present on the Affymetrix U74 microarrays, they are fairly specific, given that the QTL intervals each contain around 300 genes.

In the chromosome 1 QTL only the *Cxcr4* gene meets all four criteria, with no other gene meeting more than two (see methods). In the distal chromosome 6 QTL there are two candidates that meet three criteria: *Cxcl2* and *Cdca3*. For the chromosome 10 QTL three genes met two of the criteria: *integrin b2*, *thymopoietin* and *Spi-C*. For the chromosome 18 QTL the candidate genes are *FGF1* and *desmoglein-2* (2 criteria each). There are no candidates that meet more than one criterion for the other QTL.

Given that previous QTL have been mapped to similar regions as found in this study using different *in vitro* assays, our data thus provide genetic evidence for the presence of the LTC-IC and day 35 CAFC cells in the discrete population of L⁻S⁺K⁺ cells that

can be directly visualised by flow cytometry irrespective of the Ly-6 haplotype of the parental mouse strains. In addition to being phenotypically defined ($\text{Lin}^- \text{Sca-1}^+ \text{c-Kit}^+$), the genetic evidence are strong supportive evidence that they are enriched for functionally defined stem cells.

Discussion

We have combined the power of mouse genetics with the discrimination of flow cytometry to map QTL that control the frequency of specific sub-populations of lineage negative bone marrow cells separated according to c-Kit and/or Sca-1 labelling and that can be directly visualised irrespective of the Ly-6 haplotype. The QTL on chromosomes 1 (62 cM) and 18 (13.1 cM) that determine the frequency of a discrete population of $\text{L}^- \text{S}^+ \text{K}^+$ (R1) cells are the same or very similar to the QTL that regulate the frequency of LTC-IC (~50 cM; $p=0.0007$) (Muller-Sieburg and Riblet 1996), and day 35 CAFC (19 cM; LOD 3.3) (de Haan and Van Zant 1997) cells in BXD RI mice, respectively. A third significant QTL (chromosome 17) has previously been implicated in the frequency of $\text{Thy-1}^{\text{low}} \text{Sca-1}^+ \text{c-Kit}^+$ cells (Morrison et al. 2002) and contains the mouse H-2 complex, including the Class I MHC alleles. It is unclear whether the same gene(s) within the MHC complex determine the frequency of $\text{Thy-1}^{\text{low}} \text{Sca-1}^+ \text{c-Kit}^+$, $\text{L}^- \text{S}^+ \text{K}^+$ and $\text{L}^- \text{S}^- \text{K}^-$ cells (Tables 1 and 3) (Morrison et al. 2002).

Genetic loci that confer a susceptibility to spontaneous or induced haemopoietic malignancies have also been mapped in mice, and differences in the relative frequency of the target cell (target size) may be one genetically determined risk factor (recently reviewed by Jawad et al. 2007b). For example, QTL that determine the susceptibility to spontaneous pre-B and T lymphomas were both mapped to proximal chromosome 17 (Yamada et al. 1994a, 1994b), and thus implicates the H-2 complex in both the regulation of bone marrow stem and progenitor cell frequencies and a susceptibility to spontaneous lymphoid lymphomas. Similarly, a susceptibility to pristane-induced plasmacytoma QTL has been mapped to a position on chromosome 1 (54 cM) (Mock et al. 1993) that is very similar to QTL that determine the frequency of $\text{L}^- \text{S}^+ \text{K}^+$ cells (~65 cM; Table 3), the frequency of bone marrow LTC-IC cells (~50

cM)(Muller-Sieburg and Riblet 1996) and the relative proportion of B220 B lymphocytes in mouse peripheral blood (*Pbbcp1*; 62-64 cM)(Chen and Harrison 2002). Similarly a QTL that confers a susceptibility to radiation-induced AML was mapped to distal chromosome 1 (~100 cM; Boulton et al. 2003; Jawad et al. 2006), a position that is very similar to the *Scfr1* locus (~100 cM) that also determines the frequency of LTC-IC bone marrow cells (Muller-Sieburg and Riblet 1996). However, the distal chromosome 1 LTC-IC QTL was not revealed in this study, suggesting that the colonies in the LTC-IC clonogenic assay are phenotypically heterogeneous.

The largest effect L⁻S⁺K⁺ cell QTL detected in this study, which is on chromosome 10, may overlap with the suggestive L⁻S⁺K⁺ cell QTL found in van Os et al, 2006 (32.4-50.4cM, as compared to van Os peak between 51 and 70cM). In addition the distal chromosome 6 QTL has a similar location to the other suggestive QTL found in that study (45.9-56.0cM compared with van Os peak at 51cM).

Identifying candidate genes is unreliable with the large intervals found in QTL studies, but can be aided by *in silico* analysis (Flint et al. 2005). The only gene that met all of our criteria was *Cxcr4* on chromosome 1. This gene encodes a CXC cytokine receptor which is essential for the homing of haematopoietic stem cells to the vascular stem cell niche of the bone marrow (Burger and Burkle 2007). The fact that candidate genes for the other QTL include *Cxcl12*, which is the ligand for *CXCR4* and *Itgb2*, an integrin involved in HSC homing to the periphery (Chavakis et al. 2005), leads to the tempting hypothesis that genetic variation in a common HSC homing pathway may be important in determining stem cell numbers. This hypothesis must remain tentative however until confirmatory laboratory experiments have been carried out.

The genetic intervals mapped in these studies are large, and the possibility that multiple QTL map to the same interval cannot be excluded. The mapping of the same or very similar QTL in independent studies using different assay(s), biological end-points and/or different inbred mouse strains increases the biological and genetic significance of that QTL. Of all the L⁻S⁺K⁺ cell QTL we mapped (Table 3) on a CBA/H x C57BL/6 genetic background, the same or very similar QTL on chromosomes 1, 17 and 18 have been previously described in independent studies of

mouse stem/progenitor cell frequency using different assays and different genetic backgrounds (BXD RI, and [C57BL/Ka-Thy-1.1 x AKR/J]F2) (Table 1). Although the $L^{-}S^{+}K^{+}$ may not represent all the cells that have been functionally and genetically defined *in vitro*, the $L^{-}S^{+}K^{+}$ cells are enriched for some of those cells that give rise to LTC-IC and/or day 35 CAFC cells. Furthermore, this suggests that specific cellular phenotypes themselves can also be genetically defined, and that definition can be further refined. A combination of functional, phenotypic and genetic definitions of cell sub-populations will be an invaluable resource in providing evidence as to whether or not distinct phenotypic assays are in fact measuring the same cell(s), and potentially reveal evidence of relationships between apparently dissimilar phenotypes.

Acknowledgements

This work was supported by the Leukaemia Research Fund. CC was funded by an MRC Research Studentship, and GG supported by the HOPE Foundation for Cancer Research. We would like to dedicate this paper to Mark Plumb, who was the principal investigator on this project and died in May 2007.

References

- Boulton E, Cole C, Knight A, Cleary H, Snowden R, et al. (2003) Low penetrance genetic susceptibility and resistance loci implicated in the relative risk of radiation-induced acute myeloid leukaemia in mice. *Blood* 101: 2349-2354
- Breems DA, Blokland EA, Neben S, Ploemacher RE (1994) Frequency analysis of human primitive haemtopoietic stem cell subsets using a cobblestone area forming cell assay. *Leukemia* 8: 1095-1104
- Burger JA, Burkle A (2007) The CXCR4 chemokine receptor in acute and chronic leukaemia: a marrow homing receptor and potential therapeutic target. *Br J Haematol* 137:288-296

- Bystrykh L, Weersing E, Dontje B, Sutton S, Pletcher MT, et al. (2005) Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nature Genetics* 37:225-232
- Cervino AC, Gosink M, Fallahi M, Pascal B, Mader C, et al. (2006) A comprehensive mouse IBD database for the efficient localization of quantitative trait loci. *Mamm Genome* 17:565-574
- Chavakis E, Aicher A, Heeschen C, Sasaki K, Kaiser R, et al. (2005) Role of beta2-integrins for homing and neovascularization capacity of endothelial progenitor cells. *J Exp Med* 201:63-72
- Chen J, Harrison DE (2002) Quantitative trait loci regulating relative lymphocyte proportions in mouse peripheral blood. *Blood* 99: 561-566
- Christensen JL, Weissman IL (2001) Flk-2 is a marker in haemopoietic stem cell differentiation: A simple method to isolate long-term stem cells. *Proc Nat Acad Sci USA* 98:14541-14546
- De Haan G, Van Zant G (1997) Intrinsic and extrinsic control of hemopoietic stem cell numbers: Mapping of a stem cell gene. *J Exp Med* 186:529-536
- De Haan G, Szilvassy SJ, Meyerose TE, Dontje B, Grimes B, et al. (2000) Distinct functional properties of highly purified hematopoietic stem cells from mouse strains differing in stem cell numbers. *Blood* 96:1374-1379
- De Haan G, Bystrykh LV, Weersing E, Dontje B, Geiger H, et al. (2002) A genetic and genomic analysis identifies a cluster of genes associated with haemopoietic cell turnover. *Blood* 100:2056-2062
- Eppig JT, Bult CJ, Kadin JA, Richardson JE, Blake JA, et al. (2005) The Mouse Genome Database (MGD): from genes to mice—a community resource for mouse biology. *Nucleic Acids Res* 33:D471-D475
- Festing MF (1996) Origins and characteristics of inbred strains of mice. In: Lyon, M.F., Rastan, S. & Brown, S.D.M., eds. *Genetic Variants and strains of the laboratory mouse*. Vol 2. Oxford University Press, Oxford UK, pp1537-1576
- Flint J, Valdar W, Shifman S, Mott R (2005) Strategies for mapping and cloning quantitative trait genes in rodents. *Nat Rev Genet* 6:271-86
- Geiger H, True JM, de Haan G, Van Zant G (2001) Age- and stage-specific regulation patterns in the haematopoietic stem cell hierarchy. *Blood* 98: 2966-72
- Henckaerts E, Geiger H, Langer JC, Rebollo P, Van Zant G, et al. (2002) Genetically determined variation in the number of phenotypically defined haemopoietic

- progenitor and stem cells and in their response to early-acting cytokines. *Blood* 99:3947-3954
- Henckaerts E, Langer JC, Snoeck H-W (2004) Quantitative genetic variation in the hematopoietic stem and progenitor cell compartment and in life span are closely linked at multiple loci in BXD recombinant inbred mice. *Blood* 104:374-379
- Jawad M, Cole C, Zanker A, Lo P, Fitch S, Plumb M (2006) Evidence for clustered tumour suppressor gene loci on mouse chromosomes 2 and 4 in radiation-induced Acute Myeloid Leukaemia. *Int. J. Radiat. Biol.* 82:383-391
- Jawad M, Giotopoulos G, Fitch S, Cole C, Plumb M, et al. (2007a) Mouse bone marrow and peripheral blood red blood cell counts are regulated by different autosomal genetic loci. *Blood Cells Mol & Disease*, 38:69-77
- Jawad M, Giotopoulos G, Cole C, Plumb M (2007b) Target cell frequency is a genetically determined risk factor in radiation leukaemogenesis. *Br J Radiol* 80:56-62
- Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* 11:241-247
- Liang Y, Jansen M, Aronow B, Geiger H, Van Zant G (2007) The quantitative trait gene latexin influences the size of the hematopoietic stem cell population in mice. *Nature Genetics* 39:178-188
- Lynch M, Walsh, B (1998) *Genetics and analysis of quantitative traits* (Sunderland, MA: Sinauer Associates Inc).
- Mock BA, Krall MM, Dosik JK (1993) Genetic mapping of tumor susceptibility genes involved in mouse plasmacytomagenesis. *Proc Natl Acad Sci USA* 90:9499-9503
- Morrison SJ, Hemmati HD, Wandycz AM, Weissman IL (1995) The purification and characterization of fetal liver hematopoietic stem cells. *Proc Natl Acad Sci USA* 92:10302-6.
- Morrison SJ, Wandycz AM, Hemmati HD, Wright DE, Weissman IL (1997) Identification of a lineage of multipotent hematopoietic progenitors. *Development* 124:1929-1939
- Morrison SJ, Qian D, Jerabek L, Thiel BA, Park I-K, Ford PS, Kiel MJ, Schork NJ, Weissman IL, Clarke MF (2002) A genetic determinant that specifically regulates the frequency of hematopoietic stem cells. *J Immunol* 168:635-642

- Muller-Sieburg CE, Riblet R (1996) Genetic control of the frequency of haemopoietic stem cells in mice: mapping a candidate locus to chromosome 1. *J Exp Med* 183:1141-1150
- Okada S, Nakauchi H, Nagayoshi K, Nishikawa S, Miura Y, et al. (1992) In vivo and in vitro stem cell function of c-Kit- and sca-1-positive murine hematopoietic cells. *Blood* 80:3044-3050
- Osawa M, Hanada K, Hamada H, Nakauchi H (1996) Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* 273:242-245
- Peters LL, Lambert AJ, Zhang W, Churchill GA, Brugnara C, et al. (2006) Quantitative trait loci for baseline erythroid traits. *Mamm Genome* 17:298-309
- Petzer AL, Hogge DE, Lansdorp PM, Reid DS, Eaves CJ (1996) Self-renewal of primitive human hematopoietic cells (long-term-culture-initiating cells) in vitro and their expansion in defined medium. *Proc Natl Acad Sci USA* 93:1470-1474
- Ploemacher RE, van der Sluijs JP, van Beurden CAJ, Baert MRM, Chan PL (1991) Use of limiting-dilution type long-term marrow cultures in frequency analysis of marrow-repopulating and spleen colony-forming hematopoietic stem cells in the mouse. *Blood* 78: 2527-2533
- Spangrude GJ, Brooks DM (1992) Phenotypic analysis of mouse haemopoietic stem cells shows a Thy-1-negative subset. *Blood* 80:1957-1964
- Spangrude GJ, Brooks DM (1993) Mouse strain variability in the expression of the hematopoietic stem cell antigen Ly-6A/E by bone marrow cells. *Blood* 82:3327-3332
- Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 101:6062-6067
- Sutherland HJ, Lansdorp PM, Henkelman DH, Eaves AC, Eaves CJ (1990) Functional characterization of individual human hematopoietic stem cells cultured at limiting dilution on supportive marrow stromal layers. *Proc Natl Acad Sci USA* 87:3584-3588
- Uchida N, Weissman IL (1992) Searching for hematopoietic stem cells: evidence that Thy-1.1^{lo} Lin⁻ Sca-1⁺ cells are the only stem cells in C57BL/Ka-Thy-1.1 bone marrow. *J Exp Med* 175:175-184

- Valdar W, Solberg LC, Gauguier D, Burnett S, Klennerman P, et al. (2006) Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat Genet* 38:879-887
- Van Os R, Ausema A, Noach EJK, van Pelt K, Dontje BJH, et al. (2006) Identification of quantitative trait loci regulating haemopoietic parameters in B6AKRF2 mice. *Brit J Haematol* 132:80-90
- Wang S, Basten CJ, Zeng Z-B (2001-2004) Windows QTL Cartographer 2.0. Department of Statistics, North Carolina State University, Raleigh, NC. (<http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>).
- Yamada Y, Shisa H, Matsushiro H, Kamoto T, Kobayashi Y, et al. (1994a) T lymphomagenesis is determined by a dominant host gene *Thymic Lymphoma Susceptible Mouse-1 (TLSM-1)* in mouse models. *J Exp Med* 180:2155-2162
- Yamada Y, Matsushiro H, Ogawa MS, Okamoto K, Nakakuki Y, et al. (1994b) Genetic predisposition to pre-B lymphomas in SL/Kh strain mice. *Cancer Res* 54:403-407

Figure legends:

Figure 1:

Flow cytograms of Lin⁻ cells to illustrate the background defined by the 0-30 quadrant. A. Unlabelled Lin⁻ cells. B. Lin⁻ cells labelled with R-PE anti mouse c-Kit. C. Lin⁻ cells (CBA/H) labelled with FITC anti Sca-1 antibody.

Figure 2:

Direct visualization of Lin⁻ bone marrow cells labelled with anti Sca-1 and c-Kit antibodies by flow cytometry.

A-C. Representative flow cytograms of Lin⁻ bone marrow cells from (A.) CBA/H, (B.) C57BL/6, and (C.) DBA/2 mice labelled for Sca-1 and c-Kit expression/antibody binding and visualised directly by flow cytometry. **D.** Areas of the cytogram analysed in this study are schematically represented where the relative fluorescence intensity (1-30) is defined as negative (-) based on unlabelled cells, intermediate (+; 30^1 - 10^2) or high (++; $>10^2$) as shown. The discrete population of L-S⁺K⁺ cells that distinguish CBA/H and DBA/2 mice from C57BL/6 mice were quantitated by gating the R1 region.

Figure 3:

LOD plot using number of cells in R1 gated region (L-S⁺K⁺).

Plot showing LOD scores achieved by composite interval mapping for the L-S⁺K⁺ cell phenotype. Axis legend: Y axis shows LOD score and X axis is genetic position

with chromosomes indicated. Threshold lines are for suggestive linkage at LOD 2.5 and significant at 3.5.

Figure 4:

LOD plot using number of Sca-1⁺⁺ cells

Plot showing LOD scores achieved by composite interval mapping for the frequency of Lin⁻ Sca-1⁺⁺ cells (cytogram area $y > 10^2$). Axis legend: Y axis shows LOD score and X axis is genetic position with chromosomes indicated. Threshold lines are for suggestive linkage at LOD 2.5 and significant at 3.5.

Figure 5:

LOD plot using number of L⁻S⁻K⁻ cells

Plot showing LOD scores achieved by composite interval mapping for the frequency of double negative L⁻S⁻K⁻ cells (cytogram area $x < 30$, $y < 30$). Axis legend: Y axis shows LOD score and X axis is genetic position with chromosomes indicated. Threshold lines are for suggestive linkage at LOD 2.5 and significant at 3.5.

Table legends:**Table 1: Studies showing relevant HSC QTL.**

Mouse strains shows mapping cross: (BXD)RI is C57/BL6 x DBA recombinant inbreds.

Mouse strains	Chr (Peak)	Assay	Reference
(BXD)RI	1 (52.3 cM) 1 (98.6 cM)	LTC-IC	Muller-Sieberg and Riblet, 1996
(BXD)RI	18 (19 cM)	CAFC day 35	de Haan and Van Zant, 1997
(BXD)RI	1 (38 cM) 1 (100 cM) 18 (27 cM)	CAFC day 35	Geiger et al, 2001
(AKR/J x C57BL/Thy-1.1)F2	17 (19.8 cM)	L ⁻ S ⁺ K ⁺ Thy1 ^{lo}	Morrison et al., 2002
(C57BL/6 x AKR) F2	6 (51 cM) 10 (51-70 cM)	L ⁻ S ⁺ K ⁺	van Os et al, 2006
(C57BL/6 x AKR) F2	1 (86.6 cM)	CAFC day 35	van Os et al, 2006

Table 2:

Frequency of discrete populations of Lin⁻ bone marrow cells defined by Sca-1 and/or c-kit labelling

Refer to Figure 2D for quadrant description. (a): F1 are CBA/H x C57BL/6 hybrids. (b): Statistical significance in cellular frequencies comparing CBA/H and C57BL/6 mouse bone marrow analysed by Flow cytometry (only *p* values <0.06 are shown; Kruskal-Wallis Test).

	FACS Coordinates	CBA/H	C57BL/6	F1 ^a	<i>P</i> ^b
--	---------------------	-------	---------	-----------------	-----------------------

	Figure 1D	(n = 23-26)	(n = 17-18)	(n = 9)	(CBA vsC57)
Lin⁻, Sca-1⁻, c-kit⁻ (% Lin ⁻ ± SD)	Y = 1-30 X = 1-30	61 ± 8.3	82.2 ± 6.5	62.54 ± 4.3	
Lin⁻, Sca-1^{+/++}, c-kit^{-/+} (% Lin ⁻ ± SD)	Y = 10²-10⁴ X = 1-10²	1.0 ± 0.5	2.9 ± 1.1	3.68 ± 0.8	< 10⁻⁶
Lin⁻, Sca-1⁺⁺ (% Lin ⁻ ± SD)	Y = 10²-10⁴ X = 1-10⁴	2.73 ± 2.0	5.1 ± 1.3	5.2 ± 1.1	0.0002
Lin⁻, c-kit⁺⁺ (% Lin ⁻ ± SD)	Y = 1-10⁴ X = 10²-10⁴	26.3 ± 6.9	22.6 ± 4.8	19.4 ± 2.8	0.0525
Lin⁻, Sca-1⁺⁺, c-kit⁺⁺ (% Lin ⁻ ± SD)	Gated R1 (L⁻S⁺K⁺)	1.9 ± 0.7	0.5 ± 0.1	1.1 ± 0.4	0.0054

Table 3:

QTL found by composite interval mapping.

In Peak LOD column * shows suggestive QTL, ** significant QTL. CI= Confidence Interval. Mb positions based on genome build NCBI m36. Direction & Dominance shows the allele that increases the phenotype i.e. the number of cells in the FACS area, together with the mode of inheritance (Add = Additive, Rec = Recessive, Dom = Dominant). Candidate loci are shown in bold, candidate genes from *in silico* analyses in normal text.

FACS Area	Chrom.	Peak (cM)	Peak LOD	95% CI (cM)	Physical interval (Mb)	Direction & Mode	Candidate Locus/gene
17 19	18.4 46	4.6** 2.7*	11-22.6 22-Tel	25 – 43 24-Tel	CBA Add CBA Rec	HLA
15 17	37.6 16.5	3.7** 3.1*	23.2 – 42.9 10.7-22.9	55 – 75 23 - 44	BL6 Dom CBA Add	Sca-1/Ly-6 HLA
6 6 10 17 18	29.9 48 41.5 6.4 13.1	3.0* 2.6* 3.5** 3.2* 2.8*	25.6-31.1 45.9-56.0 32.4 – 50.4 Tel-16.5 7-21	51 - 74 104 - 127 66 - 91 Tel - 29 18 - 45	CBA Dom BL6 Add CBA Dom CBA Add CBA Rec	Cdca3, Cxcl12 Tmpos, Spi-C, Itgb2 Day 35 CAFC, FGF1, DSG2