REVIEW

Combining transcriptional profiling and genetic linkage analysis to uncover gene networks operating in hematopoietic stem cells and their progeny

**Alice Gerrits · Brad Dykstra · Marcel Otten ·
Leonid Bystrykh · Gerald de Haan**

Abstract Stem cells are unique in that they possess both the capacity to self-renew and thereby maintain their original pool as well as the capacity to differentiate into mature cells. In the past number of years, transcriptional profiling of enriched stem cell populations has been extensively performed in an attempt to identify a universal stem cell gene expression signature. While stem-cell-specific transcripts were identified in each case, this approach has thus far been insufficient to identify a universal group of core "stemness" genes ultimately responsible for self-renewal and multipotency. Similarly, in the hematopoietic system, comparisons of transcriptional profiles between different hematopoietic cell stages have had limited success in revealing core genes ultimately responsible for the initiation of differentiation and lineage specification. Here, we propose that the combined use of transcriptional profiling and genetic linkage analysis, an approach called "genetical genomics", can be a valuable tool to assist in the identification of genes and gene networks that specify "stemness" and cell fate decisions. We review past studies of hematopoietic cells that utilized transcriptional profiling and/or genetic linkage analysis, and discuss several potential future applications of genetical genomics.

Keywords Hematopoietic stem cells · Cell fate decisions · Transcriptional profiling · Genetic linkage analysis · Genetical genomics

A. Gerrits · B. Dykstra · M. Otten · L. Bystrykh · G. de Haan (✉)
Department of Cell Biology, Section Stem Cell Biology,
University Medical Center Groningen, University of Groningen,
Antonius Deusinglaan 1,
9713 AV Groningen, The Netherlands
e-mail: g.de.haan@med.umcg.nl

Introduction

Stem cells are defined by their capacity to self-renew and their ability to differentiate into mature cell types. Stem cells can be divided into two main categories: embryonic stem cells (ESCs) and "tissue-specific" stem cells. ESCs, which are derived from the blastocyst stage of the early embryo, are termed pluripotent because they are able to differentiate into cells of all three germ layers; ectoderm, endoderm, and mesoderm. In contrast, tissue-specific stem cells are termed multipotent because they are only able to differentiate into a limited number of more closely related cell types. Tissue-specific stem cells have now been identified in a number of fetal and adult tissues, where they play essential roles in organogenesis, tissue homeostasis, and repair.

One of the major challenges in the stem cell field has been to identify a universal "stem cell signature", that is, those genes that ultimately enable the hallmark stem cell features of self-renewal and pluripotency or multipotency. In addition, it is important to identify those genes that initiate differentiation and determine lineage specification. In this review, we discuss how these challenges might be addressed using the combinatorial approach of genetical genomics. In particular, we focus on hematopoietic stem cells (HSCs), one of the best studied stem cell systems, as a model to investigate cell fate decisions.

HSCs are rare cells in the bone marrow that both self-renew and generate differentiated blood cells. During the process of hematopoietic differentiation, the cells progressively amplify their numbers, lose their multipotency and become increasingly committed. Ultimately, HSCs are able to give rise to large numbers of cells of both myeloid (e.g. macrophages, granulocytes, erythrocytes, and megakaryocytes/platelets) and lymphoid lineages (e.g. T-cells and

B-cells) (Weissman et al. 2001). A simplified overview of hematopoiesis is depicted in Fig. 1. Although HSCs, intermediate progenitors, and most mature blood cells are genetically identical and are only a few cell divisions apart from each other, they differ tremendously in both phenotype and function.

In search of a stem cell signature

Since HSCs share certain hallmark properties with other stem cell types, it was speculated that a universal stem cell signature, consisting of a common set of genes whose concerted expression grant stem cells their unique properties, might exist. In an attempt to identify such a signature, the expression profiles of mouse ESCs, HSCs, and neural stem cells (NSCs) were compared by two groups independently (Ivanova et al. 2002; Ramalho-Santos et al. 2002). Both studies revealed similarities and differences between these cell populations, and generated a list of over 200 stem-cell-specific genes that mouse ESCs, HSCs and NSCs commonly expressed. Strikingly however, these two lists shared only six genes. Soon thereafter, a third independent expression profiling study comparing ESCs, NSCs, and retinal progenitor/stem cells was reported. Comparing these three independently generated lists of "stemness" genes, only one gene (Itga-6) was commonly identified (Fortunel et al. 2003). What could explain the lack of overlap between these lists? First, differences in methodology may be in part responsible (discussed in Burns and Zon 2002; Evsikov and Solter 2003). Second, ESCs are pluripotent, while the other stem cell types in the described comparisons are more restricted in their developmental
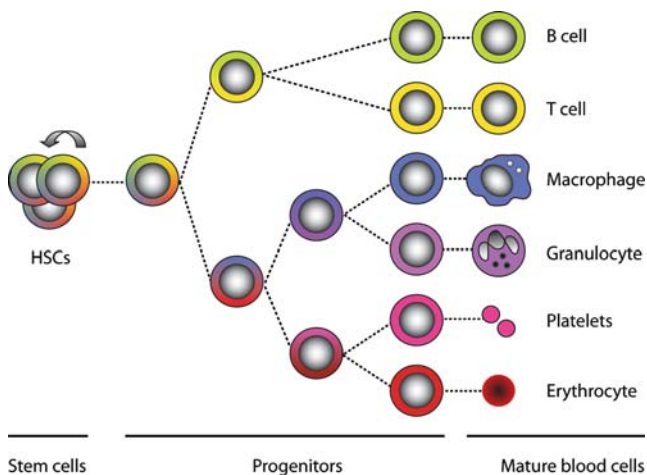
potential. These different stem cell populations were compared in a direct manner, although it is certainly possible that pluripotency and multipotency are maintained by different gene circuits. A third, more fundamental possibility is that a universal stem cell signature simply does not exist. ESCs, HSCs, NSCs, and retinal stem cells may each have their very own transcriptional networks responsible for their unique stem cell properties. If this is the case, cross-tissue comparative stem cell transcriptome analyses are not particularly useful approaches to identify these stem cell type-specific signatures.

That a limited number of factors could indeed specify "stemness" was convincingly shown by Takahashi and Yamanaka (2006), who demonstrated that the forced expression of Oct4, Sox2, Klf4 and c-Myc and could reprogram mouse fibroblasts into pluripotent stem cells that were functionally equivalent to ESCs. Subsequently, it was shown that some family proteins of the four factors could also reprogram fibroblasts, and that c-Myc was dispensable for this process (Nakagawa et al. 2008; Wernig et al. 2008). The question whether terminally differentiated cells could also be reprogrammed was resolved when fully differentiated mature mouse B lymphocytes were reprogrammed to pluripotency (Hanna et al. 2008). Induction of pluripotency was also recently demonstrated in human fibroblasts using various combinations of factors (Park et al. 2008; Takahashi et al. 2007; Yu et al. 2007). The demonstration that more than one combination of factors could confer the same stem cell characteristics suggests that stem cell signatures may be "degenerate" (as in the "*degenerate* DNA code").

Thus, it is clear that "stemness" can be achieved by only a limited number of key stem cell regulators, presumably targeting larger collections of downstream genes in a hierarchical manner. Extracting such key regulators (or causes) from their downstream target genes (consequences) is not feasible using microarray profiling approaches alone.

In search of hematopoietic fate determinants

Transcriptional profiling has also been utilized extensively in an attempt to identify genes whose expression distinguishes HSCs from their downstream progeny. Global expression analyses have revealed that stem cells exist in a "promiscuous" state where multiple lineage-specific genes are coexpressed, albeit at very low levels. Upon differentiation, "appropriate" lineage-specific genes are up-regulated, whereas "inappropriate" genes, specific for other lineages, are down-regulated (Enver and Greaves 1998). Recently, Chambers et al. generated an expression database of various hematopoietic cell types, including HSCs, erythroid cells, granulocytes, monocytes, natural killer cells, activated and naive T cells, and B cells (Chambers

Fig. 1 Simplified overview of hematopoiesis. Hematopoietic stem cells (HSCs) have self-renewal activity (represented as the *arrow*) and can therefore maintain their numbers. During hematopoietic differentiation, HSCs lose their self-renewal capacity and become increasingly lineage-committed (represented as the gradual loss of colors)

et al. 2007a). This comparative transcriptome analysis provided large lists of genes that are specifically expressed in one cell stage or cell type compared to another. However, it is improbable that the transition from one cell stage to another relies on the independent regulation of so many genes. More likely, activation of a limited number of key regulatory genes initiates a cascade of events, resulting in the altered expression of tens to hundreds of genes.

Transcriptional profiling has proven to be a useful approach to identify cell stage and cell type-specific transcripts. When combined with other genetic approaches, it may also have the potential to identify key regulatory genes.

HSCs and linkage genetics

It has become clear that many hematopoietic characteristics or traits are genetically controlled, since they differ between various strains of genetically distinct laboratory mice. For example, a substantial strain-to-strain variation in the number of primitive hematopoietic cells and their turnover rates has been observed. Interestingly, an inverse correlation was detected between progenitor cell turnover rate and mouse lifespan (De Haan et al. 1997).

Two regular inbred strains of mice, C57BL/6 (B6) and DBA/2 (D2), have distinct differences in both their HSC traits and lifespan. Compared to B6 mice, D2 mice have a shorter lifespan, a substantially higher HSC frequency, and their progenitors cycle at a much faster rate (De Haan et al. 1997; De Haan and Van Zant 1997; Muller-Sieburg and Riblet 1996; Van Zant et al. 1983). In B6 mice the HSC frequency increases at a constant rate during the aging process (Harrison et al. 1989; Liang et al. 2005; Morrison et al. 1996; Sudo et al. 2000), while in D2 mice it increases up to one year of age and then drops again (Chen et al. 2000; De Haan and Van Zant 1999a). The observed natural variation between these regular inbred mouse strains offers a powerful tool to study the genetic basis of variation in these traits. The use of B6 x D2 (BXD) recombinant inbred mouse strains has been a particularly useful strategy to identify genomic regions affecting traits of interest. These inbred lines were developed by crossing the two inbred parental strains followed by repeated sibling–sibling mating for a minimum of 20 generations. The resulting BXD mouse strains each carry a genome that consists of a unique mosaic of homozygous B6 and D2 segments. At present, the BXD panel is composed of 80 different strains that all have been fully genotyped (Peirce et al. 2004). Variation in any quantifiable trait can be associated with the segregation of parental alleles, and linkage genetics can map this variation to quantitative trait loci (QTLs), thereby identifying the genomic region(s) affecting that trait. An overview of the QTL mapping approach is depicted in Fig. 2.
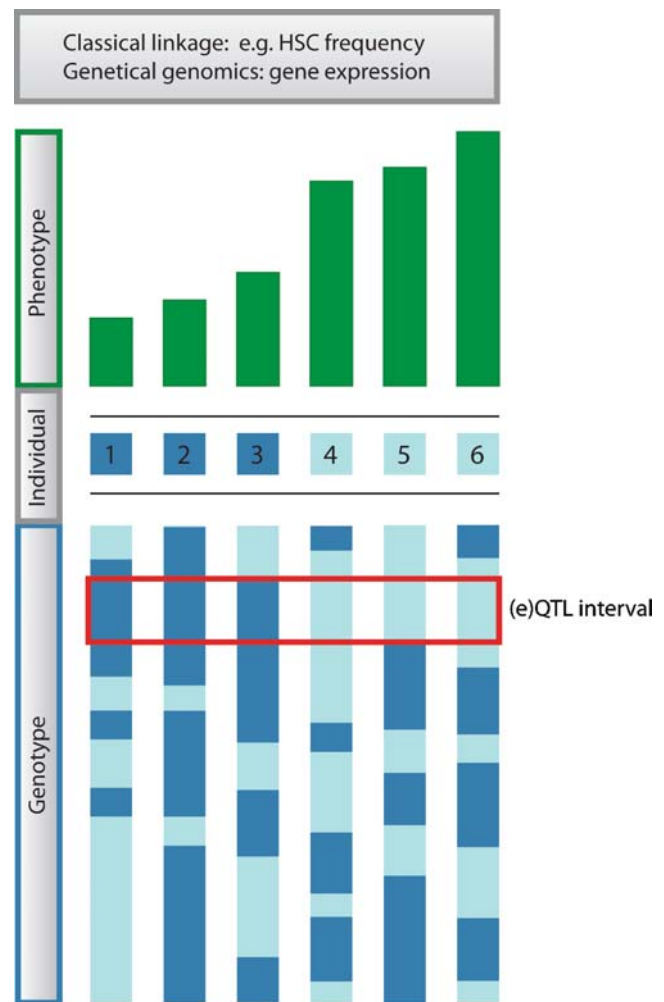

Fig. 2 Overview of (expression) quantitive trait locus mapping procedure. Variation in phenotype (here shown for six individuals) is correlated with variation in genotype (genotypes at a single chromosome are shown for each individual). The genomic location where these two parameters associate most strongly is referred to as the (expression) quantitative trait locus or (e)QTL. In this case, the three genetically distinct individuals that have a high value for the phenotype of interest carry the *light blue* genotype at the (e)QTL position, whereas the three that have a low phenotypic measure carry the *dark blue* genotype at that position. The phenotype can either be a classical trait (classical linkage) or the expression level of a gene (genetical genomics)

Classical QTL analysis has permitted the identification of loci that are associated with variation in HSC traits. When HSC frequency was measured in the BXD reference panel using long-term culture initiating cell assays (LTC-ICs), two suggestive QTLs on chromosome 1 and one on chromosome 11 were identified. One of the loci on chromosome 1 was confirmed to affect HSC frequency in a congenic mouse strain (Muller-Sieburg and Riblet 1996). When HSC frequency was assessed using cobblestone area forming cell assays, the trait mapped to a region on chromosome 18 (De Haan and Van Zant 1997). Subsequently, variation in

hematopoietic progenitor cell (HPC) frequency and HSC frequency were mapped in both young and old mice. This led to the identification of multiple QTLs, some of which were age- and differentiation stage-specific. Regardless of age, loci on chromosomes 7 and 18 were found to regulate HPC and HSC frequency, respectively. An additional locus on chromosome 1 was found to affect HPC and HSC frequency specifically in young mice, whereas loci on chromosomes 2 and 18 were found to affect these frequencies specifically in old mice (Geiger et al. 2001). A congenic mouse model was later used to confirm that the chromosome 2 locus indeed contained a regulator of HSC aging (Geiger et al. 2005). Variation in the percentage change of HSC frequency during aging was mapped to putative loci on chromosomes 2, 14, and X (De Haan and Van Zant 1999a).

Particularly interesting was the finding that variation in both turnover rate of primitive hematopoietic cells and mouse lifespan mapped to overlapping regions on chromosomes 7 and 11 (De Haan and Van Zant 1999b). This strengthened the hypothesis that mouse lifespan is in part dependent on progenitor turnover rate.

Yet another trait in which various inbred strains of mice have shown to differ is their absolute number of $Lin^- Sca-1^{++}$ cells and their responsiveness to early-acting cytokines such as kit ligand, flt3 ligand, and thrombopoietin. A genetic linkage study in BXD recombinant inbreds led to the identification of three loci on chromosomes 2, 4, and 7 that affected the total number of $Lin^- Sca-1^{++}$ cells and a locus on chromosome 2 affecting their proliferative response to cytokines. The fact that both traits mapped to the exact same region on chromosome 2 suggests that the number of $Lin^- Sca-1^{++}$ cells may depend on their responsiveness to cytokines (Henckaerts et al. 2002). This hypothesis was reinforced when a QTL for the response of primitive cells to transforming growth factor-β2 (TGF-β2) was identified on chromosome 4 that overlapped with the previously identified QTL regulating the number of $Lin^- Sca-1^{++}$ cells (Langer et al. 2004).

Furthermore, a strain-dependent variation was found to exist in the response to Granulocyte Colony-Stimulating factor (G-CSF); a growth factor that has the capacity to mobilize stem and progenitor cells from bone marrow into peripheral blood (Roberts et al. 1997). When B6 (low responder), D2 (high responder), backcross, and BXD mice were subjected to a genetic analysis, loci on chromosomes 2 and 11 (and possibly 4 and 14) were found to control G-CSF induced mobilization (Hasegawa et al. 2000).

An overview of the QTLs that have been reported to associate with various hematopoietic traits is shown in Table 1. Interestingly, multiple traits have been mapped to the same QTL regions (note the "QTL-dense" regions on chromosomes 2 and 11), suggesting that they may be regulated by a common genetic element. It should also be noted that only a few QTLs have been independently replicated. In part, this is because relatively few laboratories have used quantitative trait genetics to study hematopoiesis. Secondly, it may be due to the inherently noisy quantitative nature of the traits under study as they are likely to be controlled by multiple QTLs that each have a limited effect. Finally, multiple in vitro and in vivo assays exist that measure the functional output of primitive hematopoietic cells. Although these assays are thought to have considerable overlap with each other, they may not be measuring exactly the same spectrum of cells. Thus, it is possible that these distinct cell subsets are controlled by different genetic elements.

The major limitation of classical QTL mapping approaches in recombinant inbred reference panels is that they are only able to identify genomic *regions* of interest, usually containing tens or even hundreds of genes. This poor resolution is due to the limited number of recombination events between the two sets of parental chromosomes. Of all the genes present within the QTL interval, it is hypothesized that only polymorphic variants can be responsible for variation in the observed phenotype. In most cases, this variability is in the form of single nucleotide polymorphisms (SNPs), and to a lesser extent in the form of deletions, insertions, rearrangements, and copy number variations. Although every SNP has a potential impact on gene expression levels and therefore could affect HSC biology, the vast majority of SNPs are "synonymous" or "silent" (Kimura 1968; King and Jukes 1969). This makes it difficult to identify the causal variant or polymorphism, and therefore the causal gene that influences the trait of interest. An additional complication is the possibility that multiple linked (possibly even neighbouring) genes collectively cause the phenotype, as was proposed for the progenitor cell cycling trait (De Haan et al. 2002). A promising approach that may aid in the identification of causal genes—and the networks in which they operate—is the combination of transcriptional profiling with linkage genetics.

Transcriptional profiling and linkage genetics combined

As explained, neither transcriptional profiling alone, nor genetic linkage analysis alone, has been shown to be an effective approach to identify genes or gene networks that specify "stemness", initiate differentiation or govern lineage specification. However, the combination of both approaches may aid in their identification. Merging the fields of linkage genetics and genomics in this particular manner has been referred to as "genetical genomics" (Jansen and Nap 2001) or expression quantitative trait

Table 1 Quantitative trait loci (QTLs) associated with mouse hematopoietic traits

Trait	Chr	Region (cM)	Reference
HPC frequency	*1*	*38*	Geiger et al. 2001
HSC frequency	*1*	*38*	Geiger et al. 2001
HSC frequency (LTC-IC)	*1*	*52*	Muller-Sieburg and Riblet 1996
Splenomegaly after G-CSF	1	53–75	Roberts et al. 2000
HSC frequency (LTC-IC)	*1*	*99*	Muller-Sieburg and Riblet 1996
HSC frequency	*1*	*100*	Geiger et al. 2001
Change in HSC frequency during aging	*2*	*51*	De Haan and Van Zant 1999a
Lifespan	**2**	**69**	De Haan and Van Zant 1999b
Progenitor cell mobilization by G-CSF	**2**	**69?**	Hasegawa et al. 2000
HPC frequency (old)	**2**	**72**	Geiger et al. 2001
HSC frequency (old)	**2**	**72**	Geiger et al. 2001
No. Lin⁻ Sca-1⁺⁺ cells	**2**	**73.2–77.2**	Henckaerts et al. 2002
Response of primitive cells to cytokines	**2**	**73.2–77.2**	Henckaerts et al. 2002
HSC frequency	3	24	Geiger et al. 2001
Lifespan	4	40.3	De Haan and Van Zant 1999b
Response of primitive cells to TGF-β2	*4*	*67–79*	Langer et al. 2004
No. Lin⁻ Sca-1⁺⁺ cells	*4*	*77.5–79*	Henckaerts et al. 2002
Progenitor cell cycling	4	78	De Haan and Van Zant 1999b
Progenitor cell mobilization by G-CSF	*4*	*?*	Hasegawa et al. 2000
HSC frequency	5	20	Geiger et al. 2001
Progenitor cell cycling	7	0.5	De Haan et al. 2002
Progenitor cell cycling	*7*	*6*	De Haan and Van Zant 1999b
Lifespan	*7*	*6*	De Haan and Van Zant 1999b
HPC frequency	7	8	Geiger et al. 2001
HPC frequency	7	25	Geiger et al. 2001
HPC frequency	*7*	*50*	Geiger et al. 2001
HPC frequency (old)	*7*	*50*	Geiger et al. 2001
No. Lin⁻ Sca-1⁺⁺ cells	*7*	*51–53*	Henckaerts et al. 2002
Progenitor cell cycling	9	61	De Haan and Van Zant 1999b
Progenitor cell mobilization by G-CSF	*11*	*3*	Hasegawa et al. 2000
Progenitor cell cycling	**11**	**29**	De Haan et al. 2002
Progenitor cell cycling	**11**	**31**	De Haan and Van Zant 1999b
Lifespan	**11**	**31**	De Haan and Van Zant 1999b
HSC frequency (LTC-IC)	**11**	**32**	Muller-Sieburg and Riblet 1996
Change in HSC frequency during aging	*14*	*12*	De Haan and Van Zant 1999a
HPC frequency (old)	14	13	Geiger et al. 2001
Progenitor cell mobilization by G-CSF	*14*	*?*	Hasegawa et al. 2000
HSC frequency	15	43	De Haan and Van Zant 1997
HPC frequency (old)	15	59	Geiger et al. 2001
HSC frequency	*18*	*19*	De Haan and Van Zant 1997
HSC frequency	*18*	*27*	Geiger et al. 2001
HPC frequency (old)	*18*	*27*	Geiger et al. 2001
HSC frequency (old)	*18*	*27*	Geiger et al. 2001
Change in HSC frequency during aging	*X*	*5*	De Haan and Van Zant 1999a
HSC frequency	X	51	Geiger et al. 2001
Response of primitive cells to cytokines	X	?	Henckaerts et al. 2002

All traits were studied in C57Bl/6, DBA/2, backcross and/or BXD recombinant inbred mice. Unless noted otherwise, traits were analyzed in young mice, and HPC and HSC frequencies were measured using CAFC assays. Italic font indicates traits and corresponding QTLs that are mentioned elsewhere in this review. Entries in bold indicate "QTL-dense" regions. (Note that not all of the reported genetic associations have met the most stringent statistical threshold for significant genome-wide linkage. Also note that in the past decade, the mouse genome map has undergone significant revisions, and therfore the precise genomic locations of the identified QTLs may be slightly inacurate, especially in older publications. Nevertheless, for historical accuracy, this table shows QTL regions as specified in the original references.)

Abbreviations: *CAFC* Cobblestone area forming cell, *G-CSF* granulocyte colony-stimulating factor, *HPC* hematopoietic progenitor cell, *HSC* hematopoietic stem cell, *LTC-IC* long-term culture-initiating cell, *TGF* transforming growth factor, *?* exact region not specified

locus (eQTL) analysis (Schadt et al. 2003, 2005). The genetical genomics approach considers individual gene expression levels to be quantitative traits. In cell types isolated from genetically distinct individuals (e.g. BXD recombinant inbred mouse strains), linkage mapping can then be utilized to identify genomic regions affecting each gene expression trait (eQTL). The strategy of eQTL mapping is visualized in Fig. 2. Hence, genetical genomics studies the genetic basis of variation in gene expression. When the genomic position of the gene and the eQTL which is associated with variation in its expression level coincide, the gene is considered to be cis-regulated. In contrast, when the eQTL associated with variation in its expression level maps to another position in the genome, the gene is considered to be trans-regulated. Cis-regulation is thought to arise from either local polymorphisms in the regulatory elements that alter gene expression levels, or alternatively to arise from polymorphisms in the coding region that affect mRNA stability or feedback regulation (Fig. 3a). Cis-regulation can also originate from copy-number variability. Trans-regulation is thought to arise as a result of polymorphisms in the regulatory elements or coding region of a direct or indirect upstream regulator (Fig. 3b). It should be emphasized that in this review the terms *cis*- and *trans*-regulation do not refer to the underlying molecular nature of the regulation, but only to the distance between the physical genomic position of a gene and its corresponding eQTL position. For this reason, it is possible that a gene can be classified as cis-regulated, even though it is actually regulated by one of its neighbouring genes. To reduce the potential confusion between the type of regulation and the relative positions of genes and their eQTLs, the terms "local" and "distant" have also been proposed (Rockman and Kruglyak 2006).

Brem et al. (2002) were the first to report genetic mapping of global gene expression in a yeast cross. Since then, genetical genomics has been applied to genetically distinct strains of *Arabidopsis thaliana*, *Eucalyptus*, maize, *Caenorhabditis elegans*, mice, rats (Bystrykh et al. 2005; Chesler et al. 2005; DeCook et al. 2006; Hubner et al. 2005; Kirst et al. 2004; Li et al. 2006; Schadt et al. 2003), and also to cells isolated from human individuals (Monks et al. 2004; Morley et al. 2004). Collectively, these studies convincingly demonstrate the heritability of variation in transcript abundance and the presence of both cis- and trans-regulated genes. In addition, it is clear that the genetical genomics approach is broadly applicable to multiple species and cell types. A representative example of a genome-wide overview of genes and their eQTLs is depicted in Fig. 3c. This eQTL regulator map was created by plotting the physical genomic position of variably expressed genes and the genomic positions that were most strongly associated with variation in their expression levels

(eQTLs). Transcripts on the diagonal represent cis-regulated genes. Collections of transcripts that are identified as vertical bands or "transbands" represent genes that are located throughout the whole genome, but are thought to be transcriptionally affected by a common eQTL. If a certain genomic region harbors a higher frequency of eQTLs than expected by chance, it is termed an eQTL hotspot (Brem et al. 2002; Schadt et al. 2003). The causal regulator within such an eQTL interval may be a signalling or transcription factor that affects the expression levels of its downstream targets.

Associated with each eQTL are two important sets of genes: those regulated by the eQTL (transband genes if more than one) and those within the eQTL interval that are candidate regulators. These regulators can either directly or indirectly affect the abundance of the transband transcripts, raising the prospect of several levels of hierarchy within the regulatory network (Fig. 3d).

To identify the causal gene within each eQTL, the interval is first screened for genetic variability, since only those genes that are polymorphic can have a functional impact on the transcript abundance of the transband genes. Both regulatory elements and coding regions must be analyzed for the presence of polymorphisms. In particular, cis-acting genes are high-priority candidate regulators, as they may contain genetic variants that not only influence their own expression levels, but also those of the transband genes. Another category of genes which could affect the expression of the transband genes are those that are equally expressed, but carry polymorphisms in their coding regions, resulting in the generation of functionally distinct proteins with a differential ability to regulate downstream targets. For example, these polymorphisms can alter protein functionality by being "nonsynonymous" (amino acid changing) or alternatively by introducing alternative splicing. It should be noted that this category of candidate regulators would remain unnoticed in traditional microarray experiments, since the expression levels of such transcripts would not necessarily be altered.

While it is tempting to assume that trans-regulated genes preferentially map to eQTL intervals containing transcription factors, in yeast this could not be verified (Yvert et al. 2003). Another potential category of regulatory genes consists of signalling factors that can indirectly affect the abundance of the transband transcripts.

Although the terms cis- and trans-regulation are commonly accepted descriptions of gene regulatory relationships, their biological relevance is only assumed. While it is clear that genetical genomics has great potential to identify novel regulatory pathways and increase our understanding of regulatory networks, functional validation of candidate regulators is ultimately necessary to confirm their biological activity.
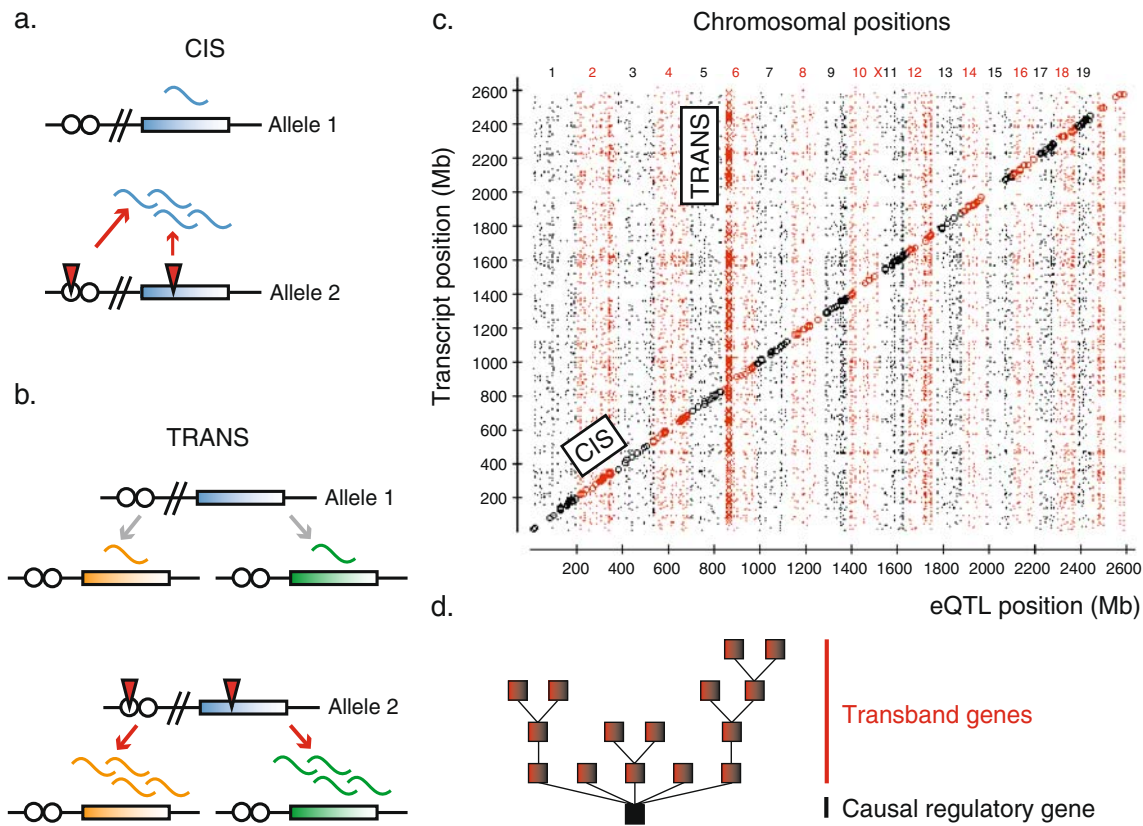
Fig. 3 Cis- and trans-regulated gene expression. Cis-regulation is expected to originate from polymorphisms (*red triangles*) in the regulatory elements (*white circles*) or the coding region (*colored rectangle*) of the gene itself (or possibly of a nearby gene) (**a**). Trans-regulation is expected to originate from polymorphisms in the regulatory elements or the coding region of a gene located distant from the gene whose expression it controls. Note that variation in expression of multiple genes can map to the same gene in trans (**b**). Transcripts and their eQTLs are graphically depicted in a genome-wide eQTL regulator map. Plotted on the y-axis are the physical positions of all measured transcripts, whereas on the x-axis the genomic regions that are most strongly associated with variation in expression levels (i.e. eQTLs) of the corresponding transcripts are shown. When transcript and eQTL position coincide, the transcript is considered to be cis-regulated and plotted on the diagonal. The vertical transband refers to transcripts encoded by genes that are positioned throughout the whole genome, but map to the same eQTL position. Transband transcripts are suggested to be coregulated. Potential transband regulators are located within the eQTL interval (where the transband meets the x-axis). Figure adapted from Bystrykh et al. 2005 (**c**). Coregulated transband genes can be directly or indirectly targeted by the potential regulator, thereby creating a network that consists of multiple levels of gene regulation (**d**)

HSCs and genetical genomics

Classical QTL mapping results in the identification of genomic intervals that affect traits of interest. This approach narrows down the number of candidate genes affecting such traits from around 30,000 (all genes in the genome) to tens or hundreds of genes (those located within that QTL interval). However, despite this major improvement, for a molecular biologist it would still require a huge effort to functionally test all the candidate genes in that interval.

A complementary approach to assist in the identification of candidate genes affecting stem cell traits is genetical genomics. Using the Affymetrix gene expression platform we collected data for each of the inbred strains from the BXD reference panel and the variation in transcript abundance in primitive Lin$^-$Sca-1$^+$c-Kit$^+$ cells of each gene was mapped to an eQTL. Using stringent significance thresholds, a total of 162 cis-regulated and 136 trans-regulated genes were identified. *Runx1*, a well-known hematopoietic regulator (Okuda et al. 1996; Wang et al. 1996), was found to be strongly cis-regulated. Interestingly, two of its known downstream targets (*Tcrb* and *Csf1r*) were found to be coregulated with *Runx1*, indicating the biological relevance of the networks that could be identified. The generation of this comprehensive eQTL dataset allowed the more detailed analysis of QTL intervals that had previously been identified using the classical approach. Cis-regulated genes within these intervals represent the best candidate regulators, as they may harbor genetic variants that

affect both their own expression levels and the trait of interest. Within the QTL interval on chromosome 11 to which we had previously mapped a region associated with hematopoietic progenitor cell turnover, only eight cis-regulated genes were identified, thereby reducing the number of candidate genes for functional testing (Bystrykh et al. 2005).

An example of the power of this combined approach was recently provided by Liang et al. Classical QTL analysis had previously identified regions on chromosomes 3, 5, and 18 that associated with variation in HSC frequency (Geiger et al. 2001). Using reciprocal congenic mouse strains, the chromosome 3 QTL interval in isolation was subsequently shown to be sufficient to confer this stem cell phenotype. Thereafter, a detailed analysis of differentially expressed transcripts within the QTL interval, followed by functional confirmation, led to the identification of *Lxn* as a gene involved in determining HSC frequency (Liang et al. 2007).

It should be noted that large collections of expression data have been deposited in the online database GeneNetwork (http://www.genenetwork.org) and are freely accessible to the research community (Chesler et al. 2004). GeneNetwork contains genotypic, phenotypic, and gene expression data from several species, including *Arabidopsis thaliana*, barley, mouse, and rat. Within the mouse BXD reference population, expression data of multiple tissues (HSCs, regulatory T cells, various neural tissues, eye, liver, lung, and kidney) are present, which permit the distinction between genes that are expressed or regulated in a tissue-specific manner from those that are equally expressed or regulated in multiple tissues.

Multidimensional genetical genomics

Genetical genomics has proven to be a valuable tool for the identification of genes and gene networks that operate in HSCs. Yet, its potential impact is only emerging and has not been fully exploited. Therefore, in this section we will address the future applications of genetical genomics.

Adding the dimension of closely related cell types

Genetical genomics has primarily been limited to single cell types. A more powerful approach would be to apply the same approach to closely related cell types. This would allow comparative analyses of gene regulatory networks between distinct but related cells. In the HSC field such multidimensional genetical genomics studies have not yet been performed. However, Li et al. (2006) recently demonstrated the general validity of the approach through the application of genetical genomics to *Caenorhabditis*

elegans recombinant inbred strains that were exposed to different temperatures. Their results showed heritable variation in gene expression responses to these environmental changes. This has created a solid basis for future multidimensional genetical genomics approaches.

In the hematopoietic system, genetical genomics could conceivably be applied to different hematopoietic cell stages (e.g. primitive HSCs, committed progenitors, and fully differentiated blood cells). Inclusion of this additional dimension would enable the study of cell fate decisions during the process of hematopoietic cell differentiation. Whereas classical analysis of gene expression levels during differentiation evaluates the dynamics of gene *expression*, analyzing eQTLs during differentiation evaluates the dynamics of gene *regulation*, permitting the identification of genes and gene networks that are specifically active in one cell type and not in another. Although eQTL profiles of different species and different cell types within the same species have previously been compared, eQTL profiles have never been evaluated for highly purified cell types that are so closely related. By implementing a "subtractive genetical genomics approach" a distinction can possibly be made between common or "housekeeping" eQTLs and those eQTLs that are specific for only one cell stage.

Multidimensional genetical genomics would also be a powerful tool to study age-dependent changes in the HSC compartment. Previous studies have compared the transcriptional profiles of HSC-enriched populations from young and old murine bone marrow. The collections of genes reported to be up-regulated in aged stem cells included those involved in inflammatory and stress responses (Chambers et al. 2007b) and signal transducer activity and receptor activity (Rossi et al. 2005), whereas those down-regulated during aging were genes involved in DNA repair and chromatin remodelling (Chambers et al. 2007b). While these studies revealed thousands of age-regulated genes, the ultimate causes of these expression perturbations remain unknown. Analyzing age-dependent gene expression changes using multidimensional genetical genomics could bring the identification of genes causing the age-induced alterations—and thereby future therapeutic intervention strategies—one step closer.

Adding the dimension of epigenetics

Epigenetic gene regulation has been suggested to play a key role in modulating stem cell fate. Epigenetics refers to heritable gene expression changes that occur without DNA sequence alterations, and includes DNA methylation and histone modifications such as acetylation, methylation, and ubiquitylation. These modifications can result in either gene activation or gene repression. That epigenetic gene repres-

sion appears to be involved in the maintenance of "stemness" became apparent when a number of developmental regulators were found to be epigenetically silenced in murine ESCs and activated upon induction of ESC differentiation (Bernstein et al. 2006; Boyer et al. 2006). It is exciting to postulate that epigenetic modifiers might similarly define cellular fate and lineage commitment during hematopoiesis.

Since a differential epigenetic conformation of the genome can result in variation in gene expression levels that can in turn affect stem cell traits, it is possible that such differential epigenetic states underlie some of the observed (e)QTL effects. At this time, however, there is insufficient knowledge on how epigenetic modifications correlate with variation in gene expression levels on a genome-wide scale. To this end, a powerful approach would be to combine traditional transcriptome profiling with whole-genome tiling arrays measuring chromatin-immunoprecipitation and DNA methylation. If these complementary array-based analyses were carried out in the same reference panel of genetically distinct individuals, not only variation in transcript abundance, but potentially also variation in epigenetic conformations could be mapped to genomic loci. In this manner, both the genetics of gene expression (i.e. "regular" genetical genomics) and the genetics of epigenetics could be studied simultaneously, thus revealing genes that directly or indirectly affect epigenetic gene states. An additional issue that could be addressed by such an approach is to estimate the percentage of variation in gene expression that can be explained by different epigenetic conformations.

The level of complexity could be further increased by including different cell types in the analysis, such as the above-mentioned different hematopoietic cell stages, different stem cell types, stem cells derived from different species, or stem cells of different ages. Through a "subtractive QTL analysis" approach, cell type specific QTLs that affect transcript abundance and/or epigenetic gene states might thus be identified. The emergence of such comparative analyses in the coming years will further the understanding of regulatory networks and how they affect cellular fate.

Adding the dimension of miRNAs

It is possible that some of the upstream regulators located within (e)QTL intervals are in fact not protein-encoding genes, but rather miRNAs. MiRNAs are small noncoding RNAs complementary to one or often multiple mRNAs, and their main function appears to be down-regulation of gene expression. Certain miRNAs have been shown to be differentially expressed between various hematopoietic cell types, suggesting that they could be involved in lineage specification (Chen et al. 2004). Polymorphisms in miRNA production sites can alter their specificity, whereas polymorphisms in regulatory elements can alter their expression levels. In addition, polymorphisms in miRNA target sites can affect the binding of a miRNA and therefore their capability to silence target gene expression. Any of these miRNA sequence variants might underlie (e)QTL effects. For example, polymorphic miRNA target sites located within a gene (often in the 3' untranslated region) might underlie a cis-acting eQTL effect. Further, transbands could be explained by a polymorphic miRNA production site that is positioned within an eQTL interval, giving rise to a miRNA that may differentially affect the expression levels of its target genes.
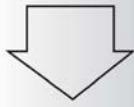
Efforts have already been made to overlap polymorphic miRNA target sites with eQTL intervals and known classical QTL intervals in an attempt to identify miRNAs that not only underlie variation in gene expression levels but also variation in cell biological traits. Genes positioned in (e)QTL intervals can be screened for polymorphisms in miRNA target sites in an online database at http://compbio.utmem.edu/miRSNP/ (Bao et al. 2007).

Little is known about the factors that regulate the expression levels of the miRNAs themselves. Genome-wide miRNA profiling of cells isolated from genetically distinct individuals would assist in this regard, since through the use of genetic linkage, variation in miRNA expression levels can be mapped to genomic regions affecting miRNA expression levels.

Adding the dimension of clinical data

The sequencing of the human genome and the development of transcriptome profiling technology have permitted new approaches to characterize hematologic malignancies at the molecular level. Gene expression profiles have been generated for malignancies such as diffuse large B-cell lymphoma, mantle cell lymphoma, acute myeloid leukemia, acute lymphoblastic leukemia, chronic lymphocytic leukemia, and multiple myeloma. All these malignancies could be classified into molecularly distinct subgroups on the basis of similarities in their gene expression profiles, and genes whose expression could discriminate between these distinct subgroups were identified (reviewed in Margalit et al. 2005). Although this subgroup classification has diagnostic, prognostic and therapeutic consequences, the disease-initiating or causative factors are still not known. If large scale clinical data and gene expression profiles were combined with detailed genotypes of the patients, this would permit the use of genetical genomics and therefore the identification of QTLs that underlie complex diseases,

Fig. 4 An overview of the described approaches (*left*), the sources of analysis (*middle*) and the phenotypic measures (*right*)

Transcriptome profiling	*One genotype* Within one or more cell types	Gene expression
Classical linkage	*Multiple genotypes* Within organism	Classical phenotype
One-dimensional genetical genomics	*Multiple genotypes* Within one cell type	Gene expression levels
Multi-dimensional genetical genomics	*Multiple genotypes* Within multiple cell types: Developmental stages Unstimulated/ stimulated Healthy/ diseased Young/ old Etc.	Gene expression levels DNA methylation status Histone modification status Protein levels Protein phosphorylation levels Etc.

and contribute to understanding which genes, gene networks and biological processes are involved in both normal and malignant hematopoietic cell development.

That genetical genomics could be applicable to human data was demonstrated in two independent studies using previously genotyped lymphoblastoid cell lines from related individuals (Monks et al. 2004; Morley et al. 2004). In both studies the heritability of gene expression levels was shown and eQTLs were identified, but there were too many differences between the two approaches to compare them in a direct manner (De Koning and Haley 2005). Association-based studies were also performed using lymphoblastoid cell lines from unrelated individuals (Cheung et al. 2005; Stranger et al. 2005). While these studies demonstrated the potential of performing genetical genomics using human data, the clinical relevance of these approaches was limited since the studies were performed on transformed human cell lines, were limited in their samples sizes, and did not include any clinical phenotypes.

A more direct example of the clinical relevance of this approach was recently provided by Goring et al. (2007), who generated genome-wide transcriptional profiles of normal untransformed lymphocytes from a large collection of genotyped individuals whose plasma cholesterol concentrations were also measured. Using a genetical genomics approach, *VNN1* was identified as a gene affecting high-density lipoprotein cholesterol concentrations.

In a second clinically relevant example, blood and adipose tissues were collected from a large group of subjects, after which genotypes, gene expression levels and clinical traits related to obesity were analyzed in a combinatorial fashion. A substantial correlation was found between gene expression profiles of adipose tissue and obesity-related traits, but not between blood expression

profiles and those traits. Importantly, genes and gene networks that were enriched in inflammatory and immune response pathways were identified that in part contribute to obesity in humans (Emilsson et al. 2008). These examples demonstrate how a multidimensional genetical genomics approach can aid in the understanding of human health.

Closing remarks

In this review, we describe how transcriptional profiling has helped to define the molecular identity of HSCs and other cell types, and how the use of linkage genetics has permitted the identification of specific genomic regions that affect HSC traits. In particular, we focus on the "genetical genomics" approach of combining transcriptional profiling with genetic linkage analysis, and discuss the potential added value of including additional dimensions in the analysis. All approaches are summarized in Fig. 4. In the coming years, multidimensional genetical genomics has the potential to greatly aid in revealing regulatory networks that specify cell fate decisions not only in HSCs, but in a whole range of clinically relevant cell types.

References

Bao L, Zhou M, Wu L, Lu L, Goldowitz D, Williams RW et al (2007) PolymiRTS Database: linking polymorphisms in microRNA target sites with complex traits. Nucleic Acids Res 35: D51–D54

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J et al (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125:315–326

Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI et al (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441:349–353

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science 296:752–755

Burns CE, Zon LI (2002) Portrait of a stem cell. Dev Cell 3:612–613

Bystrykh L, Weersing E, Dontje B, Sutton S, Pletcher MT, Wiltshire T et al (2005) Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. Nat Genet 37:225–232

Chambers SM, Boles NC, Lin KYK, Tierney MP, Bowman TV, Bradfute SB et al (2007a) Hematopoietic fingerprints: an expression database of stem cells and their progeny. Cell Stem Cell 1:578–591

Chambers SM, Shaw CA, Gatza C, Fisk CJ, Donehower LA, Goodell MA (2007b) Aging hematopoietic stem cells decline in function and exhibit epigenetic dysregulation. PLoS Biol 5:e201

Chen J, Astle CM, Harrison DE (2000) Genetic regulation of primitive hematopoietic stem cell senescence. Exp Hematol 28:442–450

Chen CZ, Li L, Lodish HF, Bartel DP (2004) MicroRNAs modulate hematopoietic lineage differentiation. Science 303:83–86

Chesler EJ, Lu L, Wang J, Williams RW, Manly KF (2004) WebQTL: rapid exploratory analysis of gene expression and genetic networks for brain and behavior. Nat Neurosci 7:485–486

Chesler EJ, Lu L, Shou S, Qu Y, Gu J, Wang J et al (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nat Genet 37:233–242

Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT (2005) Mapping determinants of human gene expression by regional and genome-wide association. Nature 437:1365–1369

DeCook R, Lall S, Nettleton D, Howell SH (2006) Genetic regulation of gene expression during shoot development in Arabidopsis. Genetics 172:1155–1164

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, Van Zant G (1999a) Dynamic changes in mouse hematopoietic stem cell numbers during aging. Blood 93:3294–3301

De Haan G, Van Zant G (1999b) Genetic analysis of hemopoietic cell cycling in mice suggests its involvement in organismal life span. FASEB J 13:707–713

De Haan G, Nijhof W, Van Zant G (1997) Mouse strain-dependent changes in frequency and proliferation of hematopoietic stem cells during aging: correlation between lifespan and cycling activity. Blood 89:1543–1550

De Haan G, Bystrykh LV, Weersing E, Dontje B, Geiger H, Ivanova N et al (2002) A genetic and genomic analysis identifies a cluster of genes associated with hematopoietic cell turnover. Blood 100:2056–2062

De Koning DJ, Haley CS (2005) Genetical genomics in humans and model organisms. Trends Genet 21:377–381

Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J et al (2008) Genetics of gene expression and its effect on disease. Nature 452:423–428

Enver T, Greaves M (1998) Loops, lineage, and leukemia. Cell 94:9–12

Evsikov AV, Solter D (2003) Comment on " 'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature" (I). Science 302:393

Fortunel NO, Otu HH, Ng HH, Chen J, Mu X, Chevassut T et al (2003) Comment on " 'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature" (II). Science 302:393

Geiger H, True JM, De Haan G, Van Zant G (2001) Age- and stage-specific regulation patterns in the hematopoietic stem cell hierarchy. Blood 98:2966–2972

Geiger H, Rennebeck G, Van Zant G (2005) Regulation of hematopoietic stem cell aging in vivo by a distinct genetic element. Proc Natl Acad Sci USA 102:5102–5107

Goring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA et al (2007) Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nat Genet 39:1208–1216

Hanna J, Markoulaki S, Schorderet P, Carey BW, Beard C, Wernig M et al (2008) Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. Cell 133:250–264

Harrison DE, Astle CM, Stone M (1989) Numbers and functions of transplantable primitive immunohematopoietic stem cells. Effects of age. J Immunol 142:3833–3840

Hasegawa M, Baldwin TM, Metcalf D, Foote SJ (2000) Progenitor cell mobilization by granulocyte colony-stimulating factor controlled by loci on chromosomes 2 and 11. Blood 95:1872–1874

Henckaerts E, Geiger H, Langer JC, Rebollo P, Van Zant G, Snoeck HW (2002) Genetically determined variation in the number of phenotypically defined hematopoietic progenitor and stem cells and in their response to early-acting cytokines. Blood 99:3947–3954

Hubner N, Wallace CA, Zimdahl H, Petretto E, Schulz H, Maciver F et al (2005) Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. Nat Genet 37:243–253

Ivanova NB, Dimos JT, Schaniel C, Hackney JA, Moore KA, Lemischka IR (2002) A stem cell molecular signature. Science 298:601–604

Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. Trends Genet 17:388–391

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624–626

King JL, Jukes TH (1969) Non-Darwinian evolution. Science 164:788–798

Kirst M, Myburg AA, De Leon JP, Kirst ME, Scott J, Sederoff R (2004) Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus. Plant Physiol 135:2368–2378

Langer JC, Henckaerts E, Orenstein J, Snoeck HW (2004) Quantitative trait analysis reveals transforming growth factor-beta2 as a positive regulator of early hematopoietic progenitor and stem cell function. J Exp Med 199:5–14

Li Y, Alvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JA et al (2006) Mapping determinants of gene expression plasticity by genetical genomics in C. elegans. PLoS Genet 2:e222

Liang Y, Van Zant G, Szilvassy SJ (2005) Effects of aging on the homing and engraftment of murine hematopoietic stem and progenitor cells. Blood 106:1479–1487

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. Nat Genet 39:178–188

Margalit O, Somech R, Amariglio N, Rechavi G (2005) Microarray-based gene expression profiling of hematologic malignancies: basic concepts and clinical applications. Blood Rev 19:223–234

Monks SA, Leonardson A, Zhu H, Cundiff P, Pietrusiak P, Edwards S et al (2004) Genetic inheritance of gene expression in human cell lines. Am J Hum Genet 75:1094–1105

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS et al (2004) Genetic analysis of genome-wide variation in human gene expression. Nature 430:743–747

Morrison SJ, Wandycz AM, Akashi K, Globerson A, Weissman IL (1996) The aging of hematopoietic stem cells. Nat Med 2:1011–1016

Muller-Sieburg CE, Riblet R (1996) Genetic control of the frequency of hematopoietic stem cells in mice: mapping of a candidate locus to chromosome 1. J Exp Med 183:1141–1150

Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T et al (2008) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol 26:101–106

Okuda T, Van Deursen J, Hiebert SW, Grosveld G, Downing JR (1996) AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell 84:321–330

Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA et al (2008) Reprogramming of human somatic cells to pluripotency with defined factors. Nature 451:141–146

Peirce JL, Lu L, Gu J, Silver LM, Williams RW (2004) A new set of BXD recombinant inbred lines from advanced intercross populations in mice. BMC Genet 5:7

Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC, Melton DA (2002) "Stemness": transcriptional profiling of embryonic and adult stem cells. Science 298:597–600

Roberts AW, Foote S, Alexander WS, Scott C, Robb L, Metcalf D (1997) Genetic influences determining progenitor cell mobilization and leukocytosis induced by granulocyte colony-stimulating factor. Blood 89:2736–2744

Roberts AW, Hasegawa M, Metcalf D, Foote SJ (2000) Identification of a genetic locus modulating splenomegaly induced by granulocyte colony-stimulating factor in mice. Leukemia 14:657–661

Rockman MV, Kruglyak L (2006) Genetics of global gene expression. Nat Rev Genet 7:862–872

Rossi DJ, Bryder D, Zahn JM, Ahlenius H, Sonu R, Wagers AJ et al (2005) Cell intrinsic alterations underlie hematopoietic stem cell aging. Proc Natl Acad Sci USA 102:9194–9199

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V et al (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297–302

Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D et al (2005) An integrative genomics approach to infer causal associations between gene expression and disease. Nat Genet 37:710–717

Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R et al (2005) Genome-wide associations of gene expression variation in humans. PLoS Genet 1:e78

Sudo K, Ema H, Morita Y, Nakauchi H (2000) Age-associated characteristics of murine hematopoietic stem cells. J Exp Med 192:1273–1280

Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126:663–676

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K et al (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131:861–872

Van Zant G, Eldridge PW, Behringer RR, Dewey MJ (1983) Genetic control of hematopoietic kinetics revealed by analyses of allophenic mice and stem cell suicide. Cell 35:639–645

Wang Q, Stacy T, Binder M, Marin-Padilla M, Sharpe AH, Speck NA (1996) Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. Proc Natl Acad Sci USA 93:3444–3449

Weissman IL, Anderson DJ, Gage F (2001) Stem and progenitor cells: origins, phenotypes, lineage commitments, and transdifferentiations. Annu Rev Cell Dev Biol 17:387–403

Wernig M, Meissner A, Cassady JP, Jaenisch R (2008) C-Myc is dispensable for direct reprogramming of mouse fibroblasts. Cell Stem Cell 2:10–12

Yu J, Vodyanik MA, Smuga-Otto K, Ntosiewicz-Bourget J, Frane JL, Tian S et al (2007) Induced pluripotent stem cell lines derived from human somatic cells. Science 318:1917–1920

Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN et al (2003) Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat Genet 35:57–64