Exploiting natural and induced genetic variation to study hematopoiesis
Gerrits, Alice

IMPORTANT NOTE: You are advised to consult the publisher’s version (publisher’s PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher’s PDF, also known as Version of record

Publication date:
2011

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
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 V. Bystrykh and 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.
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. monocytes/ macrophages, granulocytes, erythrocytes and megakaryocytes/ platelets) and lymphoid lineages (e.g. T cells and B cells). A simplified overview of hematopoiesis is depicted in Figure 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. 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 (Itga6) was commonly identified. What could explain the lack of overlap between these lists? First, differences in methodology may be in part responsible (discussed in 5,6). 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, 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.
The question whether terminally differentiated cells could also be reprogrammed was resolved when fully differentiated mature mouse B lymphocytes were reprogrammed to pluripotency.\textsuperscript{10} Induction of pluripotency was also recently demonstrated in human fibroblasts using various combinations of factors.\textsuperscript{11-13} 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 micro-array 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 co-expressed, albeit at very low levels. Upon differentiation, “appropriate” lineage-specific genes are up-regulated, whereas “inappropriate” genes, specific for other lineages, are down-regulated.\textsuperscript{14} 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.\textsuperscript{15} 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.\textsuperscript{16}

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. In B6 mice the HSC frequency increases at a constant rate during the aging process, while in D2 mice it increases up to one year of age and then drops again. 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. 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 Figure 2.

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. When HSC frequency was assessed using cobblestone area forming cell assays (CAFCs), the trait mapped to a region on chromosome 18. 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. A congenic mouse model was later used to confirm that the chromosome 2 locus indeed contained a regulator of HSC aging. Variation in the percentage change of HSC frequency during aging was mapped to putative loci on chromosomes 2, 14 and X.

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. 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−Sca1++ 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−Sca1++ 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−Sca1++ cells may depend on their responsiveness to cytokines. 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−Sca1++ cells.

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. 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.

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”. 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 neighboring) genes collectively cause the phenotype, as was proposed for the progenitor cell cycling trait.
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” \(^{37}\) or expression quantitative trait locus (eQTL) analysis.\(^{38,39}\) 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 Figure 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 (Figure 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 (Figure 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 neighboring 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.40

Brem et al were the first to report genetic mapping of global gene expression in a yeast cross.41 Since then, genetical genomics has been applied to genetically distinct strains of Arabidopsis thaliana, Eucalyptus, maize, Caenorhabditis elegans, mice, rats38,42-47 and also to cells isolated from human individuals.48,49 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 Figure 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.38,41 The causal regulator within such an eQTL interval may be a signaling or transcription factor that affects the expression levels of its downstream targets.
### Table 1. Quantitative trait loci (QTLs) associated with mouse hematopoietic traits.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Chr</th>
<th>Region (cM)</th>
<th>Reference ID</th>
<th>Ref. #</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPC frequency</td>
<td>1</td>
<td>38</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>HSC frequency</td>
<td>1</td>
<td>38</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>HSC frequency (LTC-IC)</td>
<td>1</td>
<td>52</td>
<td>Muller-Sieburg and Riblet 1996</td>
<td>17</td>
</tr>
<tr>
<td>Splenomegaly after G-CSF</td>
<td>1</td>
<td>53-75</td>
<td>Roberts et al. 2000</td>
<td>67</td>
</tr>
<tr>
<td>HSC frequency (LTC-IC)</td>
<td>1</td>
<td>99</td>
<td>Muller-Sieburg and Riblet 1996</td>
<td>17</td>
</tr>
<tr>
<td>HSC frequency</td>
<td>1</td>
<td>100</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>Change in HSC frequency during aging</td>
<td>2</td>
<td>51</td>
<td>De Haan and Van Zant 1999</td>
<td>24</td>
</tr>
<tr>
<td>Lifespan</td>
<td>2</td>
<td>69</td>
<td>De Haan and Van Zant 1999</td>
<td>29</td>
</tr>
<tr>
<td>Progenitor cell mobilization by G-CSF</td>
<td>2</td>
<td>69?</td>
<td>Hasegawa et al. 2000</td>
<td>33</td>
</tr>
<tr>
<td>HPC frequency (old)</td>
<td>2</td>
<td>72</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>HSC frequency (old)</td>
<td>2</td>
<td>72</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>No. Lin−Sca1++ cells</td>
<td>2</td>
<td>73.2-77.2</td>
<td>Henckaerts et al. 2002</td>
<td>30</td>
</tr>
<tr>
<td>Response of primitive cells to cytokines</td>
<td>2</td>
<td>73.2-77.2</td>
<td>Henckaerts et al. 2002</td>
<td>30</td>
</tr>
<tr>
<td>HSC frequency</td>
<td>3</td>
<td>24</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>Lifespan</td>
<td>4</td>
<td>40.3</td>
<td>De Haan and Van Zant 1999</td>
<td>29</td>
</tr>
<tr>
<td>Response of primitive cells to TGF-β2</td>
<td>4</td>
<td>67-79</td>
<td>Langer et al. 2004</td>
<td>31</td>
</tr>
<tr>
<td>No. Lin−Sca1++ cells</td>
<td>4</td>
<td>77.5-79</td>
<td>Henckaerts et al. 2002</td>
<td>30</td>
</tr>
<tr>
<td>Progenitor cell cycling</td>
<td>4</td>
<td>78</td>
<td>De Haan and Van Zant 1999</td>
<td>29</td>
</tr>
<tr>
<td>Progenitor cell mobilization by G-CSF</td>
<td>4</td>
<td>?</td>
<td>Hasegawa et al. 2000</td>
<td>33</td>
</tr>
<tr>
<td>HSC frequency</td>
<td>5</td>
<td>20</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>Progenitor cell cycling</td>
<td>7</td>
<td>0.5</td>
<td>De Haan et al. 2002</td>
<td>36</td>
</tr>
<tr>
<td>Progenitor cell cycling</td>
<td>7</td>
<td>6</td>
<td>De Haan and Van Zant 1999</td>
<td>29</td>
</tr>
<tr>
<td>Lifespan</td>
<td>7</td>
<td>6</td>
<td>De Haan and Van Zant 1999</td>
<td>29</td>
</tr>
<tr>
<td>HPC frequency</td>
<td>7</td>
<td>8</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>HPC frequency</td>
<td>7</td>
<td>25</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>HPC frequency</td>
<td>7</td>
<td>50</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>HPC frequency (old)</td>
<td>7</td>
<td>50</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>No. Lin−Sca1++ cells</td>
<td>7</td>
<td>51-53</td>
<td>Henckaerts et al. 2002</td>
<td>30</td>
</tr>
<tr>
<td>Progenitor cell cycling</td>
<td>9</td>
<td>61</td>
<td>De Haan and Van Zant 1999</td>
<td>29</td>
</tr>
<tr>
<td>Progenitor cell mobilization by G-CSF</td>
<td>11</td>
<td>3</td>
<td>Hasegawa et al. 2000</td>
<td>33</td>
</tr>
<tr>
<td>Progenitor cell cycling</td>
<td>11</td>
<td>29</td>
<td>De Haan et al. 2002</td>
<td>36</td>
</tr>
<tr>
<td>Progenitor cell cycling</td>
<td>11</td>
<td>31</td>
<td>De Haan and Van Zant 1999</td>
<td>29</td>
</tr>
<tr>
<td>Lifespan</td>
<td>11</td>
<td>31</td>
<td>De Haan and Van Zant 1999</td>
<td>29</td>
</tr>
<tr>
<td>HSC frequency (LTC-IC)</td>
<td>11</td>
<td>32</td>
<td>Muller-Sieburg and Riblet 1996</td>
<td>17</td>
</tr>
<tr>
<td>Change in HSC frequency during aging</td>
<td>14</td>
<td>12</td>
<td>De Haan and Van Zant 1999</td>
<td>24</td>
</tr>
<tr>
<td>HPC frequency (old)</td>
<td>14</td>
<td>13</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>Progenitor cell mobilization by G-CSF</td>
<td>14</td>
<td>?</td>
<td>Hasegawa et al. 2000</td>
<td>33</td>
</tr>
<tr>
<td>HSC frequency</td>
<td>15</td>
<td>43</td>
<td>De Haan and Van Zant 1997</td>
<td>18</td>
</tr>
<tr>
<td>HPC frequency (old)</td>
<td>15</td>
<td>59</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>HSC frequency</td>
<td>18</td>
<td>19</td>
<td>De Haan and Van Zant 1997</td>
<td>18</td>
</tr>
<tr>
<td>HSC frequency</td>
<td>18</td>
<td>27</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>HPC frequency (old)</td>
<td>18</td>
<td>27</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>HSC frequency (old)</td>
<td>18</td>
<td>27</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>Change in HSC frequency during aging</td>
<td>X</td>
<td>5</td>
<td>De Haan and Van Zant 1999</td>
<td>24</td>
</tr>
<tr>
<td>HSC frequency</td>
<td>X</td>
<td>51</td>
<td>Geiger et al. 2001</td>
<td>27</td>
</tr>
<tr>
<td>Response of primitive cells to cytokines</td>
<td>X</td>
<td>?</td>
<td>Henckaerts et al. 2002</td>
<td>30</td>
</tr>
</tbody>
</table>

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. Boxes 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 therefore the precise genomic locations of the identified QTLs may be slightly inaccurate, 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.
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 (Figure 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 “non-synonymous” (amino acid changing) or alternatively by introducing alternative splicing. It should be noted that this category of candidate regulators would remain unnoticed in traditional micro-array 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. Another potential category of regulatory genes consists of signaling 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.

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−Sca1+cKit+ 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,\textsuperscript{51,52} was found to be strongly cis-regulated. Interestingly, two of its known downstream targets (Tcrb and Csf1r) were found to be co-regulated 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

Figure 3. Cis- and trans-regulated gene expression. (A) 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). (B) 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. (C) 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 co-regulated. Potential transband regulators are located within the eQTL interval (where the transband meets the x-axis). Figure adapted from \textsuperscript{42}. (D) Co-regulated transband genes can be directly or indirectly targeted by the potential regulator, thereby creating a network that consists of multiple levels of gene regulation.
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.42

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.27 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.53

It should be noted that large collections of expression data have been
deposited in the online database GeneNetwork (http://www.g en enetwork.org)
and are freely accessible to the research community.54 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.

Multi-dimensional 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 multi-dimensional genetical genomics
studies have not yet been performed. However, Li et al 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.47 This has created a solid basis for
future multi-dimensional 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.

Multi-dimensional 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 and signal transducer activity and receptor activity, whereas those down-regulated during aging were genes involved in DNA repair and chromatin remodelling. 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 multi-dimensional 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 repression 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. 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 microRNAs

It is possible that some of the upstream regulators located within (e)QTL intervals are in fact not protein-encoding genes, but rather microRNAs. MicroRNAs are small non-coding RNAs complementary to one or often multiple mRNAs, and their main function appears to be down-regulation of gene expression. Certain microRNAs have been shown to be differentially expressed between various hematopoietic cell types, suggesting that they could be involved in lineage specification. Polymorphisms in microRNA production sites can alter their specificity, whereas polymorphisms in regulatory elements can alter their expression levels. In addition, polymorphisms in microRNA target sites can affect the binding of a microRNA and therefore their capability to silence target gene expression. Any of these microRNA sequence variants might underlie (e) QTL effects. For example, polymorphic microRNA 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 microRNA production site that is positioned within an eQTL interval, giving rise to a microRNA that may differentially affect the expression levels of its target genes.

Efforts have already been made to overlap polymorphic microRNA target sites with eQTL intervals and known classical QTL intervals in an attempt to identify microRNAs 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 microRNA target sites in an online database at http://compbio.utmem.edu/miRSNP/.

Little is known about the factors that regulate the expression levels of the microRNAs themselves. Genome-wide microRNA profiling of cells isolated from genetically distinct individuals would assist in this regard, since through the use of genetic linkage, variation in microRNA expression levels can be mapped to genomic regions affecting microRNA 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 61). 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, 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.48,49 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.62 Association-based studies were also performed using lymphoblastoid cell lines from unrelated individuals.63,64 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 sample sizes, and did not include any clinical phenotypes.

A more direct example of the clinical relevance of this approach was recently provided by Göring et al, 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.65

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. These examples demonstrate how a multi-dimensional 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 Figure 4. In the coming years, multi-dimensional 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.

Figure 4. Overview of the described approaches (left), the sources of analysis (middle) and the phenotypic measures (right).
ACKNOWLEDGEMENTS

This work was supported by the Netherlands Genomics Initiative (Horizon, 050-71-055); the Netherlands Organization for Scientific Research (VENI, 916-86-009 to B.D.; and VICI, 918-76-601 to G.d.H.); and by the European Community (Marie Curie RTN EUrythron, MRTN-CT-2004-005499).

REFERENCES

4. Fortunel NO, Otu HH, Ng HH et al. Comment on " ‘Stemness’: transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". Science 2003;302(5644):393.
22. Sudo K, Ema H, Morita Y, Nakauchi H. Age-associated characteristics of murine...


