DNA Microarray Analysis Using Statistical Methods and Clustering: Three Case Studies.

Osnat Ravid-Amir

M.Sc. Thesis submitted to the Scientific Council of the Weizmann Institute of Science

Research conducted under the supervision of Prof. Eytan Domany

February 2003

Acknowledgements

Many people helped me in my work; I would like to take this opportunity to express my gratitude.

Prof. Eytan Domany for his endless patience and guidance throughout this work. He has been truly one of my greatest teachers and I will always be grateful for his encouragement and support.

Prof. Dan Gazit for reading the text and Dr. Hadi Aslan for his continual help in this work.

Prof. Eli Canaani for reading the manuscript and for his help and advise.

Dr. Tatiana A. Rozovskaia for her patient explanations about leukemia.

Lilach Weisz and Prof. Varda Rotter for teaching me about p53.

Prof. Camil Fuchs and Prof. Yoav Benjamini for their help in statistics.

Chen Luxenberg for his explanations about bone.

Dr. Gaddy Getz, Itai Kela, Hilah Gal, Erel Levin, Noam Shental, Dr. Himanshu Agrawal, and Assif Yitzhaky who helped throughout this time and all the members of the lab, my friends.

Amnon, my husband for the discussions, thoughts and ideas, I could have never done this without him.

Abstract

DNA microarrays are a tool for testing the expression levels of thousands of genes simultaneously. Using combined supervised statistical methods and unsupervised clustering methods, we analyzed data obtained in three studies. The first study dealt with the analysis of DNA chip data obtained from acute leukemia patients, some of them possessing a chromosomal translocation in the MLL gene. MLL gene translocations are commonly involved in acute leukemias. One of the goals of this analysis was to pinpoint genes that are associated with MLL translocations. Genes specifically associated with translocation, differentiation, or in CD10- leukemia were found. In the second project we analyzed DNA microarray data of the developing bone. Our aim was to characterize the differentiation process of the stem cells and to find the genes involved. Using a novel method combining two-way ANOVA for gene screening and subsequent clustering of the genes, new genes which are involved in the stages of bone formation were found. Finally we analyzed data from samples containing mutant p53, in order to study the gain of function effects of mutations in p53.

Contents

Chapter 1
Introduction

DNA microarrays are a platform for measuring the expression levels of thousands of genes simultaneously (see chapter 5). Analyzing such large amounts of data is a major challenge. In this work we analyzed data from three different microarray experiments using various methods of analysis.

This thesis reports on the analysis of DNA microarray data obtained in experiments on three different systems.

In the first study (chapter 2) we analyzed data obtained from acute leukemia patients, some of them possessing a translocation in the MLL gene. Chromosome translocations in the MLL gene are commonly involved in acute leukemia, and one of the goals of this analysis was to find genes which are associated with this chromosome translocation. This will perhaps help to better understand the difference in cancers with and without MLL chromosome translocations.

In the second study we analyzed DNA microarray data of the developing bone. Stem cells were induced to differentiate into bone *in vivo*, by expressing the BMP2 gene. The development into bone was studied by taking samples at different timepoints. Our goal in this project was to characterize the process of differentiation into bone. Understanding of this process may help in the development of gene therapy for bone regeneration (see chapter 3).

The third study involved the p53 tumor suppressor protein. This protein is mutated in about 50% of human cancers. In addition to loss of function, mutations in the p53 gene have been found to infer also

new functions to the protein. In this study, we tried to find genes whose expression level is affected by mutant p53 (see chapter 4).

The methods of analysis used are described in detail in chapter 5.
Analysis methods can be divided to two major types: supervised and unsupervised. In supervised analysis, prior knowledge about the division of the samples into groups is used, while in unsupervised analysis no such prior information is used, and the samples are grouped according to the genes' expression levels only. Genes are also assigned to groups whose expression levels exhibit similar variation over the samples.
The supervised part of the analysis is done by hypothesis testing. For each gene separately, a null hypothesis is tested – that it's expression values over two (known) groups of samples has been drawn from the same distribution. A p-value is obtained for each gene, which denotes the probability that the null hypothesis was erroneously rejected. Thus, if the p-value is low (usually less than 5%) we conclude that the expression levels of the gene differentiates between the groups of samples compared. However, while 5% is a good value in the case when only few genes are tested, here multiple statistical tests are preformed simultaneously. Hence, for example, when performing statistical tests on 3000 genes, 150 will receive such a p-value or less even if all 3000 were drawn at random from the same distribution. In order to control the number of such “false positives” we use the FDR method, which adjusts the p-value in a way that one can bound the average fraction of false positives in the list of genes found.
For the statistical test we use the non-prametric Ranksum and permutation tests (see methods in chapter 4), which do not assume a normal distribution of the genes expression levels. In addition, we use the two-way ANOVA test which is useful for isolating particular behaviors by testing interaction effects between two different variables (for example time and condition).

The unsupervised methods include the SPC clustering algorithm and the CTWC method. In clustering, “similar” genes or samples are grouped together. SPC is a powerful clustering algorithm developed by Prof. E. Domany and M. Blatt. A major advantage of SPC is its inherent stability measure, which is assigned to each cluster. A stable cluster is composed of genes or samples which are very close to each other and far from the rest of the genes/samples. CTWC is a powerful method for reducing noise and extracting relevant biological data. CTWC focuses on correlated subsets of genes and samples, one group at a time. The relevant subsets of genes and samples are identified by means of an iterative process, which uses, at each iteration level stable gene and sample clusters that were generated at the previous step. Although CTWC can be used with a variety of clustering algorithms, we use the CTWC with the SPC algorithm which enables us to focus on stable clusters, thus making it computationally feasible.

Chapter 2

Gene Expression Analysis of Acute Leukemia

Introduction

Hematopoiesis

Hematopoiesis is the process of generation and differentiation of blood cells. All blood cells originate from a common stem cell in the bone marrow, the hematopoietic stem cell[1] (HSC)(Alberts et al., 1994). HSC are pluripotent- they have the potential to give rise to many types of cells, in this case, all the types of blood cells. In addition, HSC have the capacity to self renew. HSC are divided into three different populations of cells: long term self renewing cells, short term self renewing cells, and multipotent progenitors which lack self renewing capacity (Figure 1)(Reya et al., 2001).

During hematopoiesis, HSC divide and differentiate to generate progenitor cells committed to the myeloid lineage or, to the lymophoid lineage (Fig1). These myeloid or lymphoid progenitors continue these processes, gradually loosing their ability to proliferate, until eventually reaching a terminally differentiated state, usually living for a few days or weeks after that (Alberts et al., 1994). Mature blood cells are classified into erythrocytes (red blood cells, not containing nuclei), leukocytes (white blood cells, containing a nucleus), and platelets.

Leukemia

Leukemia is a malignant neoplasm of hematopoietic tissue originating in the bone marrow. Leukemia is characterized by abnormal unregulated proliferation of one or more cells of the hematopoietic lineage (Brown, 1993). This abnormal behavior is due to the generation of somatic mutations which confer the affected cells a proliferative or survival advantage over the normal

[1] Stem Cells– While in some tissues differentiated cells proliferate to maintain the population, in other types of tissues cells are renewed by means of stem cells. Stem cells have the capacity to divide without limit, leading to differentiated progeny and also progeny that remains as stem cells. On the other hand, the differentiating cells undergo a limited number of divisions, finally entering a terminally differentiated state, no longer being able to proliferate. This method enables the differentiated cells to concentrate their resources on their designated tasks.

cells. The mutations include translocations, deletions and insertions that usually affect oncogenes or tumor supressors.

Leukemias are classified into lymphoid (~40% of cases), and myeloid (~60% of cases), according to the lineage of the tumor cell. In addition, 55% and 45% are acute or chronic respectively (Virginia University, 2002).

Acute leukemia is characterized by proliferation of immature cells termed blasts. Its progression is fast (weeks-months) due to the rapid proliferation of the malignant cells. The clinical symptoms of the disease are mainly due to the accumulation of the nonfunctional cancerous blasts overcrowding the bone marrow, thus disrupting normal hematopoiesis and normal levels of the other types of blood cells (such as red blood cells, functional white cells and so on). Chronic leukemia involves more mature cells and the progression is more gradual (time course of years)(Brown, 1993).

In this study we focused on acute leukemias.

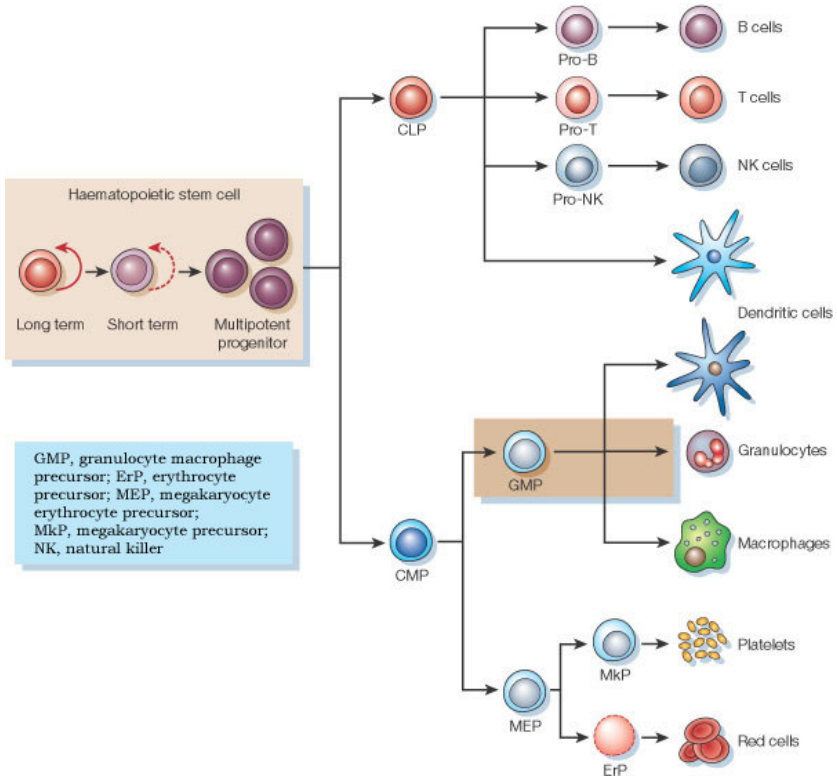

Figure 1 | Hematopoiesis: HSCs can be categorized into long-term self-renewing HSCs, short-term self-renewing HSCs and multipotent progenitors (red arrows indicate self renewal). HSC give rise to common lymphoid progenitors (CLPs) and to common myeloid progenitors (CMPs). CMPs mature into red blood cells, megakaryocyte (cells producing platelets), granulocytes, dendritic cells, and macrophages. The CLP differentiate into B and T cell lymphocytes, natural killer cells and dendritic cells. (adapted from Reya et al., 2001)

Acute Leukemia

Acute leukemia represents a clonal expansion of an abnormal immature and undifferentiated progenitor cell (Tkachuk et al., 2002). The cancerous cell may originate from a stem cell with multilineage potential or from a proginator cell developing in the myeloid or lymphoid lineages. The two main types of acute leukemia are classified according to the origin of the cancer cell: Acute Lymphoid Leukemia, ALL, originates from progenitor cells of the lymphoid pathway (that is, immature lymphocytes), and Acute Myeloid Leukemia, AML, derive from progenitor cells of the myeloid pathway (that is, cells from which neutrophils, eosinophils, monocytes, basophils, megakaryocytes, etc develop) (Fig1) (Liesner and Goldstone, 1997). ALLs and AMLs are further subdivided on the basis of morphological criteria: acute lymphoblastic leukaemia into FAB (French-American-British) subtypes L1, L2, and L3 and acute myeloid leukaemia into FAB subtypes M0 to M7 (table 1) (Liesner and Goldstone, 1997).

Table 1| FAB* classification of acute myeloid leukaemia

M0	Acute myeloid leukaemia with minimal evidence of myeloid differentiation
M1	Acute myeloblastic leukaemia without maturation
M2	Acute myeloblastic leukaemia with maturation
M3	Acute promyelocytic leukaemia
M4	Acute myelomonocytic leukaemia
M5	Acute monocytic/monoblastic leukarmia
M6	Acute erythroleukaemia
M7	Acute megakaryoblastic leukaemia

*French-American-British

Morphological classification of AML corresponds roughly to the maturation stage of myeloid, erythroid and megakaryocytic precursor cells (see table 1) The classification is usually supported by cytochemical staining.

ALL is classified into T cell lineage or B cell lineage, according to surface antigen expression. B cell lineage is further subdivided: early B precursor (pre-pre-B cell/pro B-cell) ALL is composed of the most immature cells

which do not express the common ALL antigen, CD10; cells of common ALL and pre-B cell ALL are more mature and express the CD10 antigen; B cell ALL consists of the most mature cells which express surface immunoglobulins. There is little association between morphological subtype and maturation stage and prognosis in ALL, except for the L3 morphology, which is almost exclusively found in B cell ALL (the most mature type) (Liesner and Goldstone, 1997).

AML is the most common leukemia of adulthood (Liesner and Goldstone, 1997). It constitutes approximately 80% of all acute leukemias in adults and elderly (Tkachuk et al., 2002; Bast et al., 2002), incidence rising after the age of 40 with median age of approximately 65 years (Gilliland and Tallman, 2002).

ALL occurs mainly in children and young adults and comprises 10% of all leukemias. About half of the cases of ALL are in children under the age of 15 (Leukemia-Lymphoma Society, 2002), peaking between the ages 2 and 4 (Gilliland and Tallman, 2002). In children it is the most common malignant disease and accounts for 85% of childhood leukemia (Liesner and Goldstone, 1997).

Chromosomal Abnormalities

Acquired clonal chromosomal abnormalities are detected in about 70-80% of human leukemias (Burmeister and Thiel, 2001). Most of these chromosomal abnormalities occur in a nonrandom way, and certain abnormalities are related to particular leukemia subtypes (Burmeister and Thiel, 2001). In acute leukemia, abnormalities are found in more than 60% of AML patients and in more than 65% of ALL patients (Brown, 1993). Chromosomal aberrations are present in all cases of therapy related leukemia.

The development of cytogenetic techniques (see box 1) has led to the identification of more and more recurrent chromosomal aberrations. Furthermore, as more leukemia patients underwent cytogenetic analysis, clinicians found that the chromosome abnormalities were useful prognostic indicators.

These chromosomal changes contain: gain of chromosomes, loss of chromosomes, insertions, deletions; and most prominent, chromosomal translocations (chromosome breaks and fusions).

Box 1 | **Cytogenetic analysis**

In the 1950s and 1960s, chromosomes were studied using Giemsa or Wright stains. With these techniques, the chromosomes were not banded and were counted and grouped together on the basis of similar size and shape. However, using these methods one could not distinguish between individual chromosomes in a precise way. In the 1970s, the development of chromosome banding allowed the precise identification of each chromosome and parts of chromosomes. This development led to the identification of consistently recurring aberrations in many types of cancers, including Leukemia (Rowley, 2001; Rowley, 1973a; Veldman et al., 1997).
Later on, cytogeneticists have also created fluorescence probes that hybridize to whole chromosomes, to specific portions of chromosomes or to specific genes (the method is called FISH-Florescence in situ hybridization). These probes can be used to identify certain parts of chromosomes, and to count the number of copies of each chromosome within tumour cells. A more recent technique, known as spectral karyotyping (SKY), can be used to identify individual chromosomes and rearranged chromosomes. SKY is a technology based on FISH, that is combined with another technology called spectral analysis. SKY, like FISH, employs florescent probes that attach themselves to parts of chromosomes. The tagged portion of each chromosome appears in a specific color, creating a multi-color pattern that distinguishes one chromosome from another (Rowley, 2001; Veldman et al., 1997).

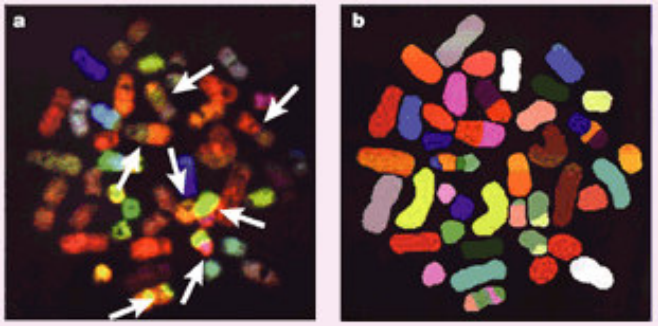

Figure 2| SKY analysis of AML cells **a**| The chromosomes are stained with a mixture of labeled probes specific for different chromosomes. Normal chromosomes show only uniform coloring, while rearranged chromosomes show two or more colors (pointed by arrows). **b**| The spectral pattern of chromosomes has been classified using computer software to identify individual chromosomes. Each chromosome has its own color code. Several colors on a single chromosome indicate a rearrangement (Rowley, 2001).

The two main outcomes of the reccurring chromosomal translocations are (fig 3): (1) juxtaposition of promoter/enhancer elements from one gene with the entire coding region of another gene, resulting in altered gene expression; (2) Recombination of coding regions of two different genes. If the reading frame is preserved, this recombination may result in a fusion protein with a novel function (Cleary, 1991).
Genes involved in translocations have very diverse functions, but it seems that most of them are involved in some critical stage of cell growth, development, or survival (Rowley, 1998). Furthermore, the affected genes are highly conserved and many are homologous to genes in other organisms (Look, 1997).

We will focus on translocation of the second type, fusion translocations, which are the most frequent type of translocations in myeloid leukemias and also compromise a large number of the lymphoid leukemia translocations (Rowley, 1998). We will discuss translocations involving the MLL gene and a translocation that results in what is known as the Philadelphia chromosome.

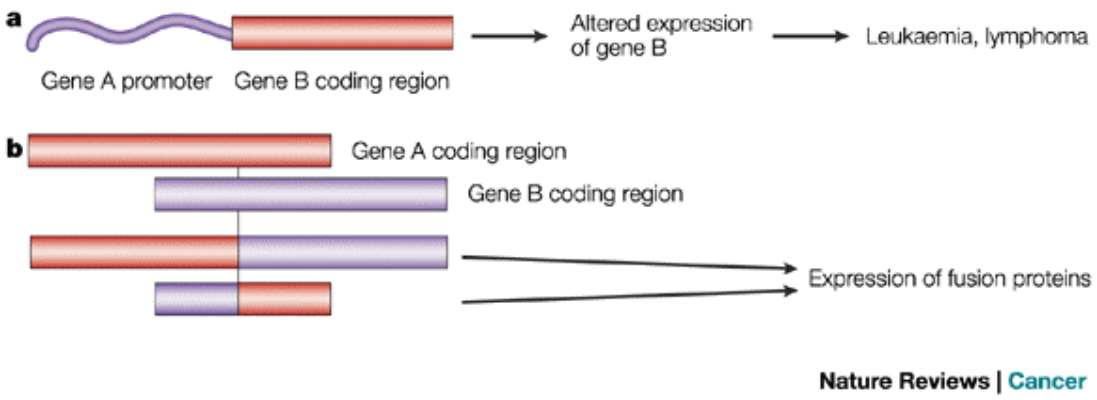

Figure 3 | **The consequences of common chromosome translocations**. a | juxtaposition of promoter/enhancer elements from one gene (gene A, purple) with the entire coding region of another gene (gene B, red), altering the normal expression of gene B. b | recombination of the coding regions of two different genes, resulting in a fusion protein that may have a novel function. (Adopted from (Rowley, 2001)).

Translocations Involving the 11q23 Chromosome Band

Translocations involving the 11q23 band are of high interest since they are implicated in 5-6% of all AML cases, 7-10% of the ALL cases, and in up to 80% of infant leukemias (<1 years of age) (Rowley, 2000; Heerema et al., 1994; Kaneko et al., 1986). Furthermore, they account for the majority of therapy related (secondary) leukemias, developing in 5-15% of primary cancer patients treated with drugs, such as etoposide (VP16) or doxorubicin (Dox) that inhibit DNA topoisomerase II (Pedersen-Bjergaard and Rowley, 1994).

Although a band such as the 11q23 band, contains 10-15 million base pairs, and therefore may contain many different genes, the MLL gene has been shown to be involved in almost all 11q23 translocations in leukemia (Rowley, 1998).

The MLL Gene

The MLL (mixed lineage leukemia, also known as ALL-1 and Htrx) (Cimino et al., 1991) gene is a large gene (100kb, 36 exons), encoding a multidomain 3968 amino acids protein. MLL is the human homologue of the Drosophila trithorax gene, which encodes a putative DNA binding protein involved in control of embryonic development (Tkachuk et al., 1992; Breen and Harte, 1993; Yu et al., 1995).

Structural Motifs: The amino-terminal region of MLL has a series of three AT hooks that bind to the minor groove of the DNA, the third hook has an ATP binding site, similar to that found in proteins which are involved in chromatin structure maintenance (Zeleznik-Le et al., 1994). MLL also has a transcriptional repression domain (Zeleznik-Le et al., 1994). Inside this domain there is a region homologous to the noncatalytic domains of DNA methyltransferase (an enzyme maintaining methylation of cytosine) (Zeleznik-Le et al., 1994). Between the AT hooks and the DNA methyltransferase homology domain there are two short conserved sequences, SNL1 and SNL2, that direct MLL to discrete subnuclear domains (Yano et al., 1997). Furthermore, a series of zinc fingers are found in a region homologous to trithorax. This zinc finger motif is similar to the plant homology domain (PHD), suggesting this motif has a role in protein-protein interactions, rather than the more usual DNA-binding role of the zinc finger domains. Finally, a transactivation domain, and a SET domain are positioned at the carboxy-terminal region of the protein. The SET domain is the region with the most similarity to the Drosophila homologue. (Tkachuk et al., 1992) (Figure 4).

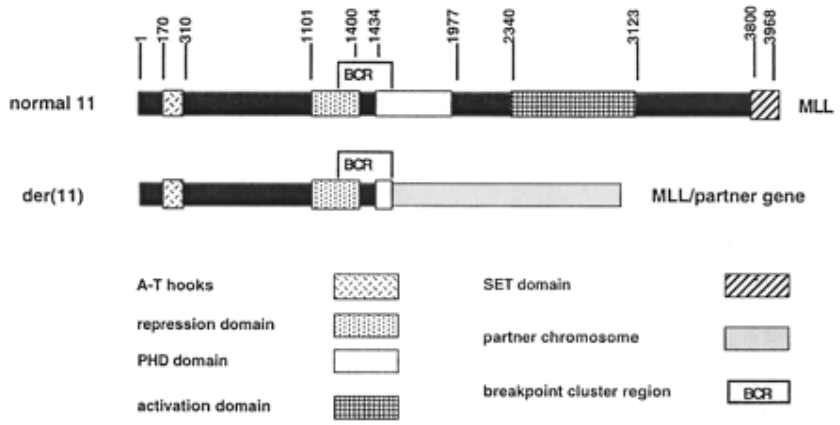

Figure 4| schematic view of the cDNA of MLL, and the cDNA of the fusion derivative der(11). The figure is drawn to scale MLL domains are identified (adopted from (Rowley, 1998))

Function: The MLL gene is highly conserved throughout evolution and is crucial for mammalian development. Indeed, mouse knockouts of MLL are embryonic lethal, leading to embryos that lack detectable expression of Homeobox (Hox) genes at day 11 in embryogenesis (Yu et al., 1995). The Hox genes are involved in embryonic development and hematopoiesis. Significantly, although in MLL null embryos correct initiation of expression of Hox genes was observed, expression was lost later (Yu et al., 1998). Thus, the embryos were not able to maintain Hox gene expression; this is consistent with the requirement for MLL Drosophila homologue TRX for the maintenance of Homeobox genes expression in flies (Breen and Harte, 1993). Furthermore, MLL knockout heterozygous mice (mice having only one active MLL allele) show skeletal abnormalities and hematopoietic cell abnormalities, all related to abnormal expression of Hox genes (Yu et al., 1995). These studies suggest that MLL plays a major role in embryonic development and hemopoietic differentiation as it regulates transcription of critical genes, such as the HOX genes (Yu et al., 1995).

Recently, the SET domain of MLL was shown to interact with INI1, a protein which is a constituent of the SWI/SNF complex, a complex acting as an ATP-dependent chromatin remodeler (Rozenblatt-Rosen et al., 1998). Moreover, two new and important studies have now shown that MLL SET

domain is a histone methyltranferase, involved in methylation of histones of type 3 on Lysine at position 4 (H3-K4, Histone 3 Lysine at position 4) (Nakamura et al., 2002; Milne et al., 2002). Generally speaking, methylation at H3-K4 is associated with transcriptional activation (Noma et al., 2001; Lachner et al., 2001; Nagy et al., 2002). The mechanism by which methylation of H3-K4 results in transcriptional activation is currently not known and is a subject of rigorous research (Noma and Grewal, 2002; Zegerman et al., 2002; Bernstein et al., 2002). Following the emergence of the "histone code hypothesis" (Strahl and Allis, 2000), it is now widely accepted that patterns of histone modifications serve as recognition code for the binding of proteins which modulate chromatin structure.

In addition to the mentioned above, these recent studies showed that MLL binds to the promoters of Hox a9 (Nakamura et al., 2002) and Hox c8 (Milne et al., 2002), mediating methylation of H3-K4 at these regions. This methylation correlates with Hox gene activation. Thus, MLL plays a direct role in transcriptional activation of Hox genes. Furthermore, Nakamura et al. showed that MLL assembles a very large supercomplex of proteins involved in chromatin modifications, RNA processing, and histone methylation. This complex remodels and covalently modifies histones, consequently regulating expression of MLL target genes. Regarding MLL fusion proteins, Milne et al. showed that although MLL fusions do not contain the SET domain, MLL-AF9 fusion protein activates Hox c8. This activation does not involve H3 Lys4 methylation.

MLL Rearrangements:

MLL is known to be involved in more than 40 different translocations to date. All characterized translocations were found to result in in-frame fusions (Rowley, 1998). The findings that all MLL fusions retain an open reading frame, indicate that this feature is strongly selected for and is likely to be indispensable for the transformation process. Until now, 17 translocation breakpoints have been cloned. All breakpoints occur within a 8.3kb region of the MLL gene, the breakpoint cluster region, BCR (Fig3). Within the fusion protein, the N-terminal ~1400 amino acids originates from

MLL, and the carboxy terminus is derived from the partner gene (Review-Rowley, 1998).

In some cases of AML, the MLL gene rearranges by internal partial duplication spanning exons 2–8, 2-6, or 4-6 (Schichman et al., 1995; Schichman et al., 1994). It has been shown that the partially duplicated MLL is transcribed into mRNA encoding a partially duplicated protein (Mecucci et al., 2002). These findings imply that the MLL protein plays a critical role in the leukemogenesis process and that internal rearrangements within the protein are sufficient to make it oncogenic.

All types of MLL fusion proteins retain the amino terminal AT hook motifs involved in DNA binding, consistent with the possibility that the fusion proteins and normal MLL bind to the same target genes. While the structural diversity of MLL partners is puzzling, it might suggest that the partners share some common function such as dimerization. However, this issue has not yet been resolved. Another issue debated over the years has been whether MLL rearrangements involve gain or loss of function. In support of a mechanism of gain of function is the well-established notion that the different partner proteins correlate with the phenotype of the leukemia (ALL or AML) as well as with clinical features. In addition, studies with transgenic mice show that MLL fusion genes act as dominant alleles (Dobson et al., 1999; Corral et al., 1996).

The possibility of MLL loss of function has also been suggested. This could involve haplo-insufficiency, meaning that the loss of one normal MLL allele is sufficient for triggering leukemia. Another scenario is that the fusion protein prevents the activity of the single normal MLL allele retained (dominant negative inhibition). Haploinsufficiency is not likely because no cases of acute leukemia were observed in which one allele of MLL was simply truncated or deleted. In addition, MLL fusion proteins do transform myeloid progenitors in the presence of two copies of w.t. MLL gene (Lavau et al., 1997). Dominant negative inhibition is not consistent with the observation that transgenic mice expressing MLL/AF-9 develop normally and appear normal for a long period of time before the onset of disease (Dobson et al., 1999). In summary, the current opinion is that MLL

rearrangements involve gain of function. This gain of function might be due to the conversion of MLL to a constitutively active transcriptional effector. In support of this is the fact that some target genes of normal MLL, such as Hoxa7 and Hoxa9 are consistently activated in leukemias with MLL translocations (Rozovskaia et al., 2001; Armstrong et al., 2002; Yeoh et al., 2002). In addition, exogenous MLL-AF9 was shown to bind the Hoxc8 promoter and concurrently activate expression of Hoxc8 (Milne et al., 2002).

Partner Genes:

Among the 17 cloned partner genes there are no obvious common structural features which could elucidate their role in the leukmogenesis process. However, MLL fusion proteins seem to fall into two functional categories: signaling molecules that are normally localized to the cytoplasm/cell junctions, or nuclear factors implicated in different aspects of transcriptional regulation (Ayton and Cleary, 2001). Furthermore, some of the partner genes share similar domains. Examples are AF4 on chromosome 4, AF9 on chromosome 9, and ENL on chromosome 19, involved in the t(4:11), t(9:11), and t(11:19) translocations, respectively. The three proteins contain nuclear localization signals as well as proline rich regions which may function as transactivation domains (Nakamura et al., 1993). Another relevant example are AF10 and AF17, both providing a leucine zipper dimerization domain to the MLL/AF10 and MLL/AF17 fusion proteins, respectively.

In this context, homo-dimerization of the amino terminus of MLL was shown to be sufficient to turn MLL into a strong transcriptional activator in reporter assays (Galoian et al., 2000), as well as to make the protein leukemogenic (Dobson et al., 2000).

AF-4: a protein of 1210 amino acids. Contains many serine and proline rich sequences, a nuclear targeting sequence and a consensus sequence for ATP/GTP binding, functions as a transcription activator (Huret and Marschalek, 2002). Interestingly, transcriptional activation activity was localized to a central conserved domain, which is consistently retained in all

MLL-AF4 fusions. In addition, AF4 function is required for normal hematopoiesis (Isnard et al., 2000).
The fusion protein consists of ~1400 amino acids of MLL containing the AT hook and the DNA methyltransferase domains and 850 amino acids from the C-terminus of AF4 (variable breakpoints). Typically this fusion protein is associated with de-novo infant ALL, and with therapy-related ALL (Huret and Marschalek, 2002).

AF-6: a protein of 1612 amino acids located both at cell-cell junctions and in the nucleus. Contains a GLGF motif, which may have a role in membrane/cytoskeleton function.
The fusion protein contains the NH2-terminus of MLL and most of the AF6 protein. Associated in M4/M5 AML (Huret, 1997b).

AF-9: A 568 amino acids protein; rich in serine and proline; possesses a nuclear targeting sequence. Functions as a transcription activator (Huret, 1997c). The fusion protein contains the C-terminal 78 residues of AF9. ALL1-AF9 protein is mostly associated with M5/M4 de novo AML and therapy related AML (Huret, 1997c).

AF-10: A 1027 amino acid protein. Contains 3 zinc fingers, Glu/Lys rich domain, a leucine zipper and a poly Ser domain. Acts as a transcription factor. The fusion protein includes the leucine zipper dimerization domain. Mainly implicated in M4/M5 AML (Huret, 1997a). The suggested mechanism for transformation mediated by the fusion protein is through the leucine zipper domain (Ayton and Cleary, 2001).

ENL: A 559 amino acid serine/proline rich protein; contains a nuclear targeting sequence and a consensus sequence for ATP/GTP binding. Acts as a transcription activator. The fusion protein contains the nearly complete ENL protein (Huret, 1997d). The transactivation domain of ENL is conserved with Anc1 a component of two yeast transcription complexes SWI/SNF, and TFIIF and TFIID (Cairns et al., 1996). Thus, the ENL moiety probably

provides a transcriptional transactivation domain to the fusion protein. The fusion protein is implicated in B-ALL, AML (M4/M5) and treatment related leukemia (Huret, 1997d).

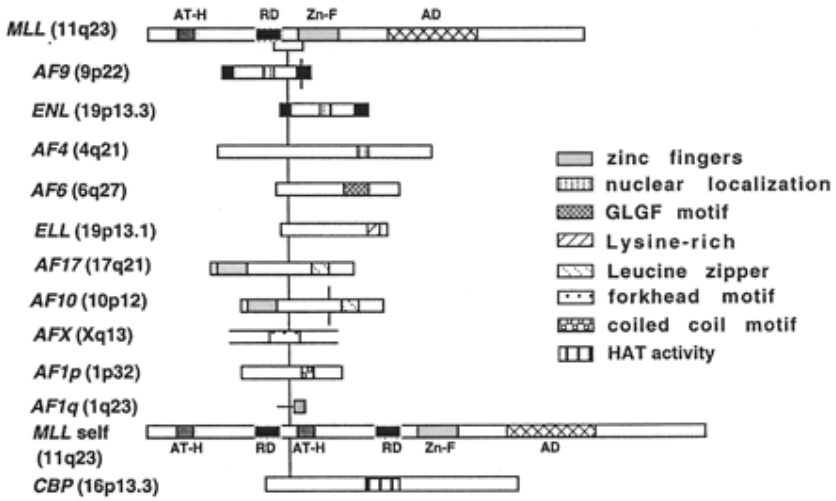

Figure 5| Motifs in some of the MLL fusion partners, and within partially duplicated MLL (self fusion). Black boxes on ENL and AF9 indicate region of homology with unknown function (adopted from (Rowley, 1998)).

The Philadelphia Chromosome

The Philadelphia chromosome results from a translocation involving chromosomes 9 and 22 (Rowley, 1973b). This small chromosome is present in all CML patients and was the first chromosomal abnormality which consistently associated with a human malignant disease. Later on, a similar translocation was found in ALL patients (Berger et al., 1990).

In both cases the two genes involved are the BCR (Breakpoint Cluster Region (Groffen et al., 1984) gene on chromosome 22, and the ABL gene on chromosome 9, creating a fusion gene composed of the 5' segment of the BCR gene and the 3' portion of the ABL gene (Canaani and Gale, 1985; Shtivelman et al., 1985; Canaani et al., 1984; Gale and Canaani, 1984; Fainstein et al., 1987; Heisterkamp et al., 1983). The ABL protein is a tyrosine kinase, which is induced by the BCR fusion partner. The chimeric polypeptides expressed from the resulting BCR-ABL oncogene form a tetramer (The bcr gene contains an oligomerization domain) that exhibits constitutive ABL kinase activity. Although ABL is normally localized to the

nucleus, addition of the BCR segment causes the BCR-ABL oncoprotein to be localized to the cytosol. The fusion protein binds to many intracellular signal-transduction proteins (proteins that ABL would not normally activate) phosporylating them, and activating signal transduction pathways, leading to uncontrolled cell growth (Harvey et al., 2002).

It was shown that the presence of the BCR-ABL transcript is sufficient and necessary for the development of CML (Daley et al., 1990). Later on, a new drug, STI-571, a kinase inhibitor was developed on the basis of all this knowledge and revolutionized CML therapy (Druker et al., 2001).

Our goal was to study the gene expression profiles of acute leukemias and to understand the role of MLL translocations in the disease.

Recent Leukemia Microarray Studies

Korsmeyer And Golub et al, Nature Genetics January 2002 (Armstrong et al., 2002)

Korsmeyer and Golub et al used DNA chips to show that ALL with MLL translocations have a distinct gene expression profile and can be distinguished as a unique subtype of leukemia separable from both ALL with no MLL translocation, and AML.
The data was analyzed using supervised statistical analysis and PCA. In addition to that, a three-class predictor was developed to classify leukemia samples into one of the three classes: ALL with no MLL translocation, AML and ALL with MLL translocation.

Statistical Analysis: Gene expression profiles of B-precursor ALL bearing an *MLL* translocation (17 samples) were compared against those from individuals diagnosed with conventional B-precursor ALL that lack this translocation (20 samples). The statistic that was used was a signal to noise statistic:

$$test\ \ statistic = (m_{ALL} - m_{MLL}) \Big/ (s_{ALL} + s_{MLL})$$

Where m and s are the median and standard deviation of the expression level of one gene, respectively, for each class. The subscript ALL denotes "conventional" ALL, without MLL translocation, MLL denotes ALL bearing the MLL translocation.
In order to estimate the p-values corresponding to the measured statistic, a hundred permutations of the samples were carried out, to estimate the chance to obtain such a statistic value or better by chance. Genes having a p-value smaller than 1% were declared as significant. K et al identified 1000

genes relatively down expressed in MLL samples and 200 genes relatively highly expressed.

In addition to that the same analysis was preformed in order to distinguish each known class of the samples: AML, conventional ALL (no translocation), and MLL, from the other two classes. The analysis found about 200 genes that are significantly overexpressed in MLL as compared with the other two leukemia categories (No more findings were published).

A drawback of this analysis is that it does not account for the multiplicity problem (see methods section for detailed explanation of the multiplicity problem). Therefore, some of the genes that were found as differentially expressed are due to statistical fluctuations rather than to a real effect.

PCA: PCA is a technique used in order to transform the data from multi-dimensional space into a lower dimensional one, that can be visualized. In this type of analysis, linear combinations of variables are ordered according to their ability to explain the variability in the data, such that the first components explain most of the variability in the data; these are the first principal components. The top three principal components of two data sets were presented (figure 6): (a) PCA of the full dataset of 8,700 genes that passed filtering. (b) PCA using a dataset of the top 500 genes that correlated with AML/ALL class distinction.

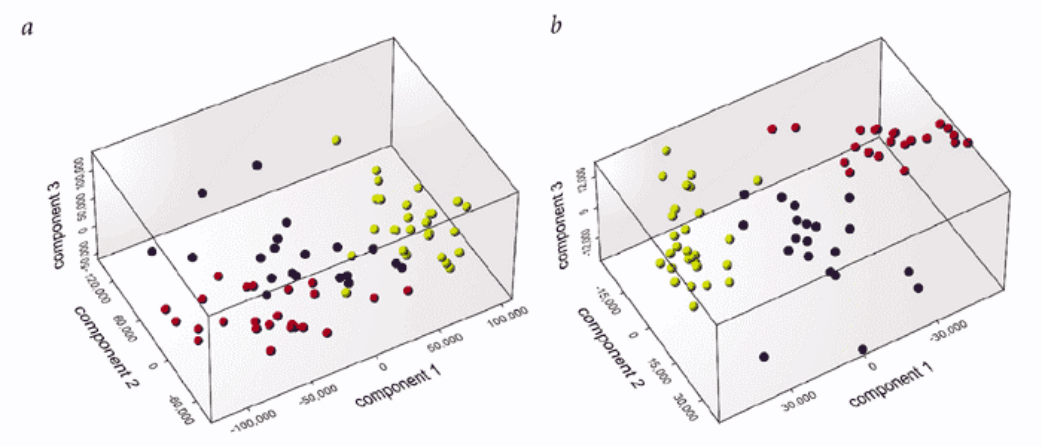

Figure 6| Comparison of gene expression between ALL, MLL and AML. ***a***, Principal component analysis (PCA) plot of ALL (red), MLL (blue) and AML (yellow) carried out using 8,700 genes that passed filtering. ***b***, PCA plot comparing ALL (red), MLL (blue) and AML (yellow) using the 500 genes that best distinguished ALL from AML.

Prediction Algorithm: A three-class predictor based on a *k*-nearest-neighbors algorithm was developed. This algorithm assigns a test sample to a class by identifying the *k* nearest samples (using Euclidean distance) in the training set and choosing the most common class among these *k* nearest neighbors. The training set for the model consisted of 20 ALL, 17 MLL and 20 AML samples. In order to assess the model a cross validation method was utilized: each time, 1 out of the 57 samples was removed, the genes that most closely correlated with the ALL/MLL/AML class distinction were identified and the expression of these genes was used to establish the class of the suspended sample. The model allocated the withheld sample to the appropriate class with 95% accuracy. Furthermore, when classifying a test set consisting of ten independent samples ten out of ten samples were correctly classified.

Conclusions: The main point the authors make is that ALL with t(4:11) translocations are a distinct class among the acute leukemias.

Downing et al Cancer Cell, March 2002 (Yeoh et al., 2002).

Downing et al. analyzed DNA microarray data of 360 pediatric ALL patients, and demonstrated that using the microarray platform patients can be accurately classified to the relevant leukemia subgroup, such that the overall diagnostic accuracy is 96%.
Downing et al. analyzed the data using various methods including clustering analysis, statistical analysis, and Multidimensional Scaling. Classification of samples was carried out using a supervised learning algorithm.

Samples consisted of 6 known leukemia subtypes including T-ALL, E2-PBX1, BCR-ABL, TEL-AML1, MLL rearrangement, hyperdiploid>50 chromosomes, and, in addition to that, a heterogeneous group of leukemias which were not categorized.

Clustering Analysis: Two dimentional hierarchical clustering (using Pearson correlation coefficient and unweighted pair group method using arithmetic averages, GeneMath 1.5) identified the 6 classified subtype groups of leukemia and a new group of samples with a distinct gene expression profile. This group consisted of 14 cases and did not have any *consistent* cytogenetic abnormality. No further information about these samples was given in the article.

Statistical Analysis: Statistical analysis was used in order to identify genes that are differentially expressed in each subgroup (including the newly identified group of leukemias) with respect to the other subgroups. The statistic that was used was the Chi-square statistic. This test tests each gene individually by measuring the Chi-square statistics with respect to the classes. The genes are then sorted according to their Chi-square statistics. The higher the X^2, the further the gene's distribution of expression values from what one would expect to find by chance in random data, and thus the more significant is the gene.

Classification: The most striking result of Downing et al. was the achieved accuracy of their class prediction using the microarray platform. Using 2/3 (215 samples) of the samples as a training set, the supervised learning algorithm learnt, through an iterative process of error minimization, to recognize the optimal gene expression pattern of each subtype.
Classification was preformed using the supervised learning algorithm with a set of discriminating genes that were chosen in one of a few possible ways: correlation base feature selection-CFS, X^2 test as described, T-test, etc.
The accuracy of the model was then tested on a test set consisting of a third of the samples (112 samples), and reached 96%. This accuracy stayed approximately the same when discriminating genes were chosen in different ways.
A similar attempt was preformed trying to predict relapse. No unifying expression signature for relapse was found for all leukemia subtypes. However, within specific subtypes of leukemia, distinct expression profiles

were identified that could predict relapse. Yet, these results proved statistically significant only for T-ALL and hyperploid>50 leukemia.

Conclusions: Using gene expression profiling, the authors demonstrated that a single platform of microarray expression analysis can accurately identify each of the known classes of childhood ALL.

Results and Discussion

This work was done in collaboration with Prof. Canaani and Dr. Tatiana A. Rozovskaia from the Molecular Cell Biology Dept. at Weizmann Institute.

Samples

The samples were provided by the GIMENA Italian Multicenter Study Group and collected from both ALL and AML patients (Table 16 in Appendix A).
The ALL samples consisted of 3 T-cell ALLs; 6 Pre-B cell ALLs, three of them containing a PH chromosome; 4 CD10- pre-pre-B cell ALLs (we will refer to them as CD10-); and 14 t(4:11) ALLs (ALLs with a fusion of chromosome 4 to chromosome 11, referred to as MLL), 2 of which were cell lines.
The AML samples consisted of 10 AML samples at M4 or M5 stages; 3 AML samples with MLL partial duplication; 9 AML samples with t(9:11), 4 of which were cell lines; 1 t(10:11) AML, 1 t(11:19) AML, and 1 sample from an AML cell line with t(6:11).
In addition, there were 8 *Normal* samples from healthy individuals.

Preprocessing and filtering

The experiments were done using two versions of the U95A chip. The expression data of the tumor samples were organized in a matrix of n_S=*52* columns (experiments) and 12,600 rows (genes on the chip). First, we removed non-human Affymetrix controls and genes appearing on only one of the versions of the U95A chip. The remaining data were preprocessed and filtered as described in the methods section using a threshold value of T=10 for the thresholding operation and $\sigma_T = 1.1$ for the filtering operation.
3064 genes passed all these operations. All further analysis was done using these genes only.

Supervised Analysis

We use the Wilcoxon Rank Sum test (Mann Whitney test) (see Methods chapter 5) in order to find genes differentially expressed between two groups

of samples. At first we focused on the ALLs, later we turned to analyze the AMLs as well.

MLLs Compared to ALLs Without an MLL translocation.

In this analysis we were seeking genes that are differently expressed in ALLs bearing a translocation and ALLs without a translocation. A Ranksum test was preformed comparing the 14 MLL (ALLs with MLL trans) samples with the other 13 ALLs (3 T-cell ALL, 6 pre-B-cell ALL and 4 pre-pre-B Cell ALL, all with no translocations, referred to as CD10-). At an FDR level of 5% we found about 240 genes that separate these two groups (full list of the genes can be found in table 1 of Appendix A) (Figure 7). However, the two groups differ not only in their translocation status, but also in their differentiation stages; as the MLL are pre-pre-B cell leukemias and most of the samples of the other group were at later stages. Hence we could not conclude that the separating ability of the genes that we found can be unequivocally attributed to the effect of the chromosome translocation.

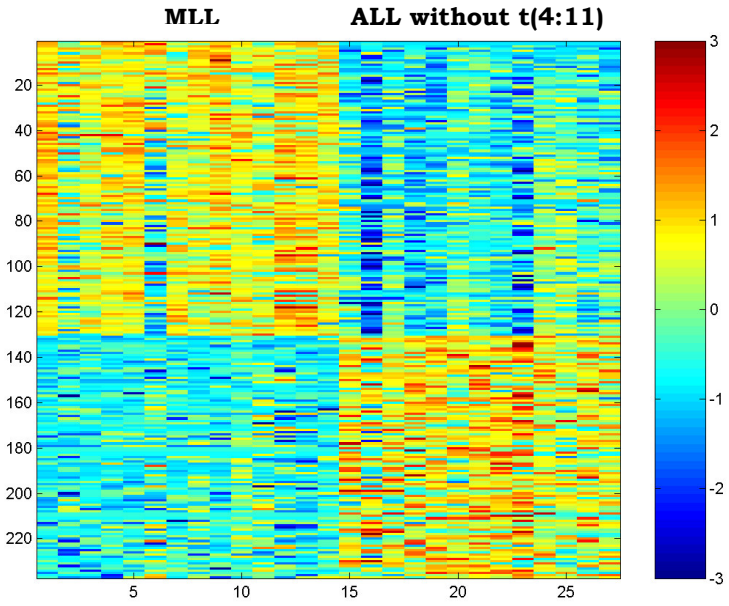

Figure 7 | Genes differentially expressed between ALL with t(4:11) and ALL without. The Values are shown as standard deviations from the mean.

The finding of so many genes separating MLLs from other ALLs at FDR 5% is remarkable. This shows that MLL and ALL without t(4:11) have very

different expression profiles, further supporting the conclusion drawn by Korsmeyer et al. that MLL can be classified as a distinct class of leukemia.

Finding Genes Specifically associated with t(4:11) tumors, CD10- ALLs and Differentiation

In order to specifically pinpoint genes associated with t(4:11) ALLs (MLLs), we preformed another analysis. The lymphocytic leukemias were divided into three groups: (i) MLLs, (ii) CD10- and (iii) pre-B cell and T cell ALLs (we refer to these as ALL). As evident from the following table every pair of groups share **one** common attribute. Hence a gene that differentiates members of group X from Y, cannot be sensitive to their shared attribute.

Attribute / Group	**t(4:11) Translocation***	**CD10- Tumors†**	**Non-Early Differentiation‡**
MLL	Yes	No	No
CD10-	No	Yes	No
ALL	No	No	Yes

***t(4:11) chromosome translocation and associated features**

†Features specific to CD10- tumors

‡Non-Early Differentiation stage

Three Ranksum tests were preformed, comparing two groups at a time (the FDR thresholds were set to 5% for MLL vs ALL and 12% for the other two comparisons).

We found 448 genes that separate MLLs (group i) from ALLs (group iii). These are expected to be associated with ALLs carrying the chromosome translocation and/or in distinguishing differentiation stages, but definitely **not** in a response specific to the CD10- tumors (group ii). On the other hand, when comparing MLLs to CD10- ALLs (group ii), we found 144 genes, expected to be linked with the t(4:11) translocation and/or having expression levels specific to CD10- tumors, but **not** sensitive to differentiation stage (since both groups are pre-pre B cell ALLs). Therefore the 46 genes from the intersection of these two tests, i.e. those that separate (i) from (iii) **and** (i) from (ii), are specific **neither** to the CD10- cells **nor** to

differentiation; therefore these genes are most likely affected by the t(4:11) translocations and/or by unique features of the cell in which the translocation occured (Table 1, figure 8 shows the intersections). From now on we will refer to these genes as translocation related genes.

In a similar way, the 167 genes that separate CD10- tumors (group ii) from ALLs (group iii) are expected to be either involved in differentiation stage or in specific features of the CD10- tumors, but should **not** be sensitive to the chromosome translocation and associated features. Therefore, their intersection with those genes that separate MLL (group i) from ALL (group iii), will include genes likely to be involved directly in hematopoietic differentiation or to distinguish between pre-pre B-cell tumors and pre B and T-cell tumors; 80 such genes were found (list of genes can be found in Appendix A table 5). From now on we will refer to these genes as differentiation genes.

Finally, the intersection of the comparison between ALL and CD10- tumors with the comparison between MLL and CD10- ALLs, results in partition of genes specific to CD10- ALLs; 21 such genes were found (Figure 8, full list of genes is given in Appendix A table 6). From now on we will refer to these genes as CD10- genes.

The expression levels of the three groups of genes derived from the three intersections, are shown in figure 9. Each group of genes splits into upregulated and downregulated subgroups[2].

Inspection of figure. 9 points to three t(4;11) tumors - samples 2, 6, 14, which show variant transcription profiles. While the expression pattern of genes unique to the translocation is similar in these three tumors and in the rest of t(4;11) ALLs, the transcription profile of the three tumors with regard to the 80 genes associated with differentiation stage is closer to T-cell and Pre-B-cell ALLs. These three tumors also appear to vary from the other t(4;11) ALLs in transcription of the genes specifically associated with CD10- ALLs.

[2] There were three additional genes that were found in the intersections, which were excluded, these are discussed in Appendix B.

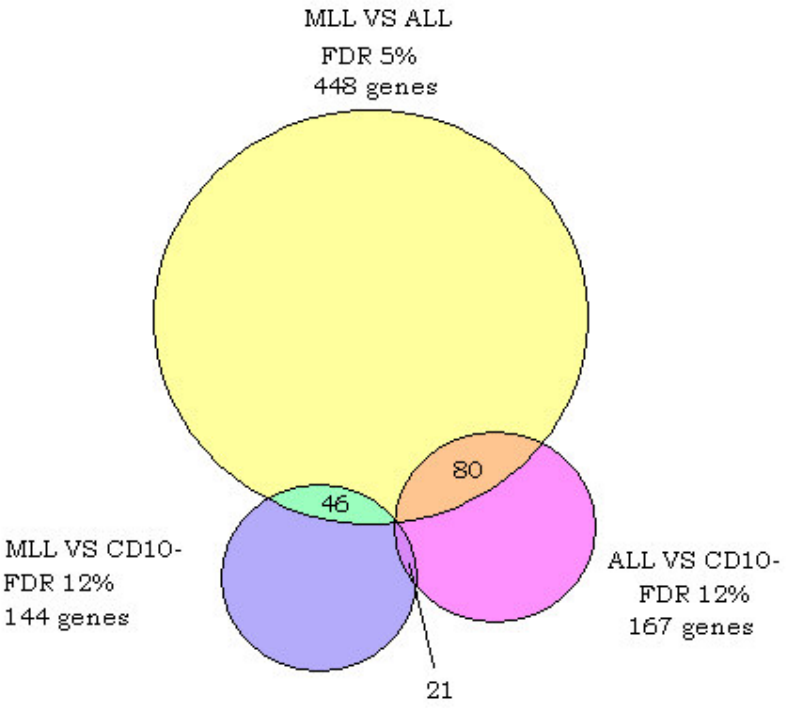

Figure 8 | Intersections of the genes found in three different comparisons. MLL, denotes MLL translocations; ALL; pre-B cell and T-cell ALLs; CD10-, pre-pre-B cell ALLs without translocations. The orange intersection contains 80 genes that are associated with differentiation, the green intersection contains 46 genes that are associated with MLL translocation, and the purple intersection contains 21 genes that are associated with CD10- pre-pre-B cell leukemia. The triple intersection was empty except for one gene which was excluded because it separated the CD10- samples from ALL and MLL in different directions (see Appendix B).

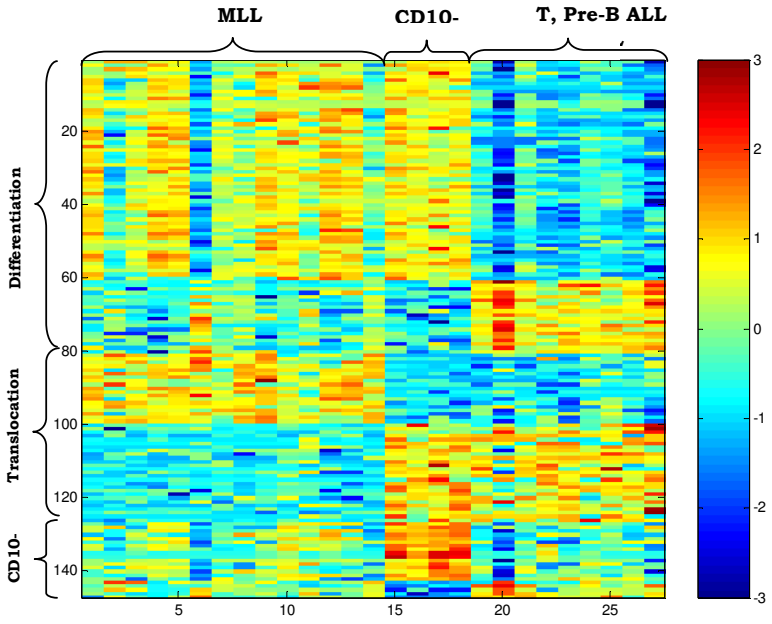

Figure 9 | Expression values of the genes found in the different intersections. The Values are shown as standard deviations from the mean.Gene 1 to 80 genes are the genes that were found to be differentially expressed between pre-B cell and T-cell ALL and MLL and also between pre-B cell and T-cell ALL and CD10- pre-pre-B cell ALL. Genes 81 to 126 are the genes that were found to be differentially expressed between MLL and pre-B cell and T-cell ALL and also between MLL and CD10- pre-pre-B cell ALL. Genes 127 to 147 are genes that were found to be differentially expressed between CD10- pre-pre-B cell ALL and pre-B cell and T-cell ALL and also between CD10- pre-pre-B cell ALL and MLL.

Although a variant expression pattern of samples 2 and 6 was already noticed before (Figure 7), only the last analysis established the root of the variance. Thus, the different expression pattern of tumor 2, 6 and 14 does not involve genes whose expression correlate with unique features of t(4:11) ALLs, but rather involves the differentiation genes.

These findings suggest that there exist two subfamilies of ALLs with t(4:11). One possibility is that the two subfamilies differ in differentiation stage.

Genes Found to be Associated With Unique Features of t(4:11) Tumors

Among the 46 genes identified by this analysis as t(4:11) associated there are many connected with growth control, cell transformation or malignancy. Those genes may be classified into several functional categories:

Upregulated in MLL

Oncogenes such as Hox A9, Meis, Hox A10, MYC, and LGALS1 are overexpressed in MLLs . Hox A9 and Meis1, form a sequence specific DNA binding complex (Shen et al., 1999), and are frequently co-activated in spontaneous AML of BXH-2 mice (Nakamura et al., 1996). Forced co-expression of the two genes in murine bone marrow cells rapidly induces AML (Kroon et al., 1998). Hox A10, was found to induce AML in mice (Thorsteinsdottir et al., 1997). MYC has a critical role in cell proliferation and is deregulated in human lymphomas and other tumors (Nesbit et al., 1999). LGALS1 (galectin1), overexpression correlates with progression of glioblastoma (Yamaoka et al., 2000), and inhibits T-cell survival and proliferation (Rabinovich et al., 2002).

In addition to these, CD44 a gene associated with drug resistance, CD72, which is involved in the proliferation and differentiation of B cells, and PPP2R5C a gene involved in survival and protection from apoptosis are also upregulated. CD44 is associated with aggressive B-CLL (Eistere et al., 1996) and confers resistance to several widely used anticancer drugs (Fujita et al., 2002). PPP2R5C (phosphatase 2A), is implicated in regulation of growth, transcription and signal transduction. Furthermore PPP2R5C is required for survival and protects from apoptosis in Drosophila (Li et al., 2002).

Downregulated in MLL

Pro-apoptotic genes such as ITPR3 (inositol 1,4,5-triphosphate receptor type 3), and JUN are underexpressed in MLLs compared to other ALLs. ITPR3 mediates the release of intracellular calcium and consequently actively promotes apoptosis (Blackshaw et al., 2000). JUN was implicated as a positive modulator of apoptosis induced in hematopoeitic progenitor cells of the myeloid linkage (Liebermann et al., 1998). Downregulation of JUN might account for the failure of glucocorticoid therapy (Pallardy and Biola, 1998).

Several tumor suppressors and growth inhibitors are also underexpressed including FHIT (fragile histidine triad), DAPK1 (death-associated protein kinase 1) and MADH1 (mothers against decapentaplegic homologue 1; SMAD1). FHIT is a target of chromosome aberrations and is inactivated in

many cancers, including lung, oesophagus, stomach, breast, kidney and leukemias (Pekarsky et al., 2002). DAPK1 counters oncogene-induced transformation by activating a p19ARF/p53 apoptotic checkpoint, and is induced by TGF-beta (Raveh et al., 2001; Jang et al., 2002). MADH1 (mothers aginst decapentaplegic homologue 1; SMAD1) is transcription modulator mutated in various forms of cancer (Hata et al., 1998).

Interestingly, there are some genes involved in proliferation and progression of cell cycle which are underexpressed in t(4:11) ALLs compared to other ALLs. These genes include CD69 which is involved in lymphocyte proliferation, and CCND2 whose overexpression leads to cell-cycle progression. In addition, FLT1 and FLT3LG are slightly overexpressed in ALLs compared to t(4:11) ALLs, both genes are implicated in hematopoietic cell proliferation.

Table 1| t(4:11)-Associated Genes

No	Acc. No	Symbol and name
OVEREXPRESSED in MLL		
1	M14087	HL14, beta-galactoside-binding lectin
2	L40377	SERPINB8, serine (or cystein) proteinase inhibitor, clade B (ovalbumin), member 8
3	AF072099	immunoglobulin-like transcript 3 protein, variant 1
4	M54992	CD72 antigen; B cell proliferation and differentiation
5	U41813	HOXA9, homeo box A9; sequence specific transcription factor, murine myeloid leukemia
6	M58597	FUT4, fucosyltransferase 4 (alpha (1,3) fucosyltransferase, myeloid-specific); glycosylation
7	AC004080	HOXA10, homeo box A10; sequence specific transcription factor
8	AA099265	RECK, reversion-inducing-cystein-rich protein with kazal motifs; membrane glycoprotein, tumor suppression
9	AF041248	CDKN2C, cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4); negative control of cell proliferation
10	D16532	VLDLR, very low density lipoprotein receptor; signal transduction, modulation of Dab1/Tau phosphorylation
11	D78177	QPRT, quinolinate phosphoribosyltransferase; biosynthesis of NAD and NADP
12	U85707	MEIS1, myeloid ecotropic viral integration site 1 homolog (mouse); homeobox protein, murine leukemia
13	Z69030	PPP2R5C, protein phosphatase 2, regulatory subunit B (B56), gamma isoform
14	AI535946	LGALS1, lectin, galactoside-binding, soluble 1 (galectin1); cell apoptosis and differentiation
15	L06419	PLOD, procollagen-lysine, 2-oxoglutarate 5-dioxygenase (lysine hydroxylase, Ehlers-Danlos syndrome)
16	V00568	MYC, v-myc myelocytomatosis viral oncogene homolog (avian); transcription factor, cell proliferation
17	AF070523	JWA, tumor rejection antigen (gp96) pseudogene 1
18	M59040	CD44, CD44 antigen (Indian blood group system); cell surface receptor, lymphocyte activation/homing
19	L05424	CD44, CD44 antigen (Indian blood group system); cell surface receptor, lymphocyte activation/homing
UNDEREXPRESSED in MLL		
20	L20433	POU4F1, POU domain class 4, transcription factor1; synaptogenesis, axonogenesis
21	S80864	CYCL, cytochrome c-like antigen; tumor antigen

22	AB007895	KIAA0435
23	AB002301	KIAA0303
24	AL050173	C21orf25, chromosome 21open reading frame 25
25	AF084481	WFS1, Wolfram syndrome 1 (wolframin)
26	U59912	MADH1, mothers against decapentaplegic homolog 1 (Drosophila); receptor-regulated transcription
27	Z22576	CD69, CD69 antigen (p60, early T-cell activation antigen); receptor, lymphocyte proliferation
28	M92357	TNFAIP2, tumor necrosis factor, alpha-induced protein 2
29	U70321	TNFRSF14, tumor necrosis factor receptor superfamily, member 14; lymphocyte activation
30	X76104	DAPK1, death-associated protein kinase 1; mediating interferon-gamma-induced cell death
31	X58072	GATA3, GATA-binding protein 3; transcriptional activator of T cell receptor alpha and delta genes
32	U19969	TCF8, transcription factor 8; repression of transcription (interleukin-2)
33	D13639	CCND2, cyclin D2; cell cycle control
34	U01062	ITPR3, inositol 1,4,5-triphosphate receptor type 3; signal transduction, small molecule transport
35	S80562	CNN3, calponin 3, acidic; thin filament-associated protein, smooth muscle contraction
36	X53586	ITGA6, integrin alpha 6; laminin receptor, critical structure role in the hemidesmosome
37	U46922	FHIT, fragile histidine triad; nucleotide metabolism, tumor suppressor
38	U46922	FHIT, fragile histidine triad; nucleotide metabolism, tumor suppressor
39	U01134	FLT1, fms-related tyrosine kinase 1, vascular endothelial growth factor receptor
40	J04111	JUN, v-jun sarcoma virus 17 oncogene homolog (avian); transcription factor
41	L34059	CDH4, cadherin 4, type 1, R-cadherin (retinal); cell adhesion
42	U59423	MADH1, mothers against decapentaplegic homolog 1 (Drosophila); receptor-regulated transcription
43	M16594	GSTA2, glutathione S-transferase A2
44	U03858	FLT3LG, fms-related tyrosine kinase 3 ligand; stimulates proliferation of early hematopoietic cells
45	S59184	RYK, RYK receptor-like tyrosine kinase
46	J03600	ALOX5, arachidonate 5- lipoxygenase; biosythesis of leukotrienes

AMLs without MLL translocation Compared to AMLs with translocation

AMLs with MLL translocations (9:11, 6:11, 11:19, and 10:11, total of 12 samples) were compared in their expression profiles to AMLs with no translocations (10 samples). At FDR of 0.1 we identified 75 genes which were differentially expressed between AML with MLL translocations and AMLs with no translocations.

Genes Associated with MLL translocations in AML

Examination of the list of genes most correlated with AMLs carrying MLL chromosome translocations (Table 2) discloses some involved in cancer or related processes. These include the overexpressed insulin receptor which enhances DNA synthesis and inhibits apoptosis (Tseng et al., 2002), the overexpressed repair gene RAD 51 which is upregulated in breast and pancreatic cancers (Maacke et al., 2000) and probably increases drug

resistance, the overexpressed PPP2R5C phosphatase, the underexpressed JUNB which upregulates the tumor suppressor gene p16 and represses cyclin D1 (Shaulian and Karin, 2001) and whose knockout in mice induces myeloproliferative disease (Passegue et al., 2001), the underexpressed tumor suppressor FHIT, the underexpressed double stranded RNA-activated protein kinase proapoptotic PRKR, which upregulates FAS and BAX (Gil and Esteban, 2000), and the underexpressed DEFA1 (defensin) involved in immune response.

Table 2 | Genes Differentially Expressed in AML with MLL rearrangement compared to the other AMLs. FDR =0.1* †

OVEREXPRESSED in AML with MLL translocation

No	Gene ID	Acc. No	Symbol and function	p-value
1	35318_at	AB007944	KIAA0475	0.000171
2	41431_at	AB023153	ICK, intestinal cell kinase	0.000213
3	33912_at	Y13834	ZMPSTE24, zinc metalloproteinase (STE24 homolog, yeast)	0.000222
4	41260_at	U59321	DDX17, DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 17 (72 kD); helicase, RNA-dependent ATPase	0.000318
5	33162_at	X02160	INSR, insulin receptor; signal transduction	0.000325
6	34936_at	AB012130	SLC4A7, solute carrier family 4, sodium bicarbonate cotransporter, member 7	0.000369
7	34737_at	AF058718	COG5, component of oligomeric golgi complex 5	0.00037
8	37303_at	AF057160	ADPRTL1, ADP-ribosyltransferase (NAD+; poly (ADP-ribose) polymerase) -like 1	0.00037
9	32044_at	D13635	KIAA0010, ubiquitin-protein isopeptide ligase (E3)	0.000475
10	706_at	B93370	NR3C1, nuclear receptor subfamily 3, group C, member 1; regulation of gene expression	0.000606
11	37973_at	AB018256	SNX13, sorting nexin 13	0.000636
12	41480_at	AF029669	RAD51C, RAD51 homolog C (S.cerevisae)	0.000758
13	38353_at	AF042378	GCP3, spindle pole body protein; microtubule cytoskeleton	0.000771
14	1944_f_at	AF001359	MLH1, DNA mismatch repair protein, alternatively spliced	0.000928
15	36048_at	AB015342	HRIHFB2436; endocrine regulator	0.000978
16	40437_at	AL049944	DKFZp564G2022	0.000978
17	36498_at	AI936759	AP1S1, adaptor-related protein complex 1, sigma 1 subunit; clathrin coat assembly protein	0.00105
18	39166_s_at	D83174	SERPINH2, serine/cysteine proteinase inhibitor, clade H (heat shock protein 47), member 2; collagen synthesis	0.00105
19	40982_at	AA926957	FLJ10534	0.00123
20	1348_s_at	S79219	PCCA, propionyl Coenzyme A carboxylase, alpha polypeptide; fatty acids metabolism	0.00123
21	176_at	U37352	PPP2R5C, protein phosphatase 2, regulatory subunit B (B56), gamma isoform	0.00123
22	36935_at	M23379	RASA1, RAS p21 protein activator (GTPase activating protein) 1	0.001234
23	549_at	S80343	RARS, arginyl-tRNA synthetase	0.001234
24	36968_s_at	AL050353	OIP2, Opa-interacting protein 2	0.001551
25	38035_at	AF072928	MTMR6, myotubularin related protein 6	0.001551
26	41557_at	D29641	KIAA0052	0.001551

27	32081_at	AB023166	CIT, citron (rho-interacting, serine/threonine kinase 21)	0.00172
28	31869_at	AB014540	SWAP70, SWAP complex protein, 70 kD	0.001941
29	37985_at	L37747	LMNB1, lamin B1; inner nuclear envelope	0.001941
30	40086_at	D87450	KIAA0261	0.001941
31	39097_at	X63753	SON, SON DNA binding protein; sequence specific transcription factor	0.001941
32	2000_at	U26455	ATM, ataxia telangiectasia mutated; kinase family, signal transduction, cell cycle, DNA repair	0.001941
33	40732_at	D83243	NPAT, nuclear protein, ataxia telangiectasia locus	0.002156
34	39761_at	D87682	KIAA0241	0.002211
35	41248_at	AB014589	CSTF2T, likely ortholog of mouse variant polyadenylation protein CSTF-64	0.002413
36	37663_at	X70649	DDX1, DEAD/H (Asp-Glu-Ala-Asp/His) box popypeptide 1; helicase family	0.00242
37	40459_at	S69189	ACOX1, acyl-Coenzyme A oxidase 1, palmitoyl; energy pathways, lipid metabolism	0.002443
UNDEREXPRESSED in AML with MLL translocation				
38	37192_at	U28389	EPB49, erythrocyte membrane protein band 4.9 (dematin); villin/gelsolin family, actin-bundling protein	0.000171
39	1008_f_at	U50648	PRKR, protein kinase, interferon-inducible double stranded RNA dependent; inhibition of protein synthesis	0.000171
40	31506_s_at	L12691	NTR3, neutrophil peptide-3	0.000213
41	39908_at	AF069735	TAF6L, TAF6-like RNA polymerase II, p300/CBP-associated factor (PCAF)-associated factor, 65 kDa	0.000222
42	37149_s_at	U95626	LTR, lactotransferrin; iron transport	0.000241
43	293_at	X74864	HPX42, homeo box protein	0.000241
44	37405_at	U29091	SELENBP1, selenium binding protein 1	0.000252
45	32052_at	L48215	HBB, hemoglobin, beta; oxygen transport	0.000265
46	35473_at	Z74615	COL1A1, collagen, type1, alpha 1; skeletal development, epidermal differentiation	0.000286
47	31687_f_at	M25079	HBB, hemoglobin beta; oxygen transport	0.000356
48	37285_at	X60364	ALAS2, aminolevulinate, delta-, synthase 2 (sideroblastic/hypochromic anemia); heme biosynthesis	0.00037
49	35601_at	L00022	IGHE, immunoglobulin heavy constant epsilon	0.000493
50	33530_at	M33326	CEACAM8, carcinoembryonic antigen-related cell adhesion molecule 8	0.000503
51	31525_s_at	J00153	HBA1, hemoglobin, alpha 1; oxygen transport	0.000606
52	38715_at	W04490	GYPB, glycophorin B (includes Ss blood group); minor sialoglycoprotein in erythrocyte membranes	0.000606
53	32821_at	AI762213	LCN2, lipocalin 2 (oncogene 24p3)	0.000632
54	31793_at	AL036554	DEFA1, defensin, alpha 1, myeloid-related sequence	0.000771
55	35919_at	J05068	TCN1, transcobalamin I (vitamin B12 binding protein, R binder family)	0.000784
56	40062_s_at	X58851	MYL4, myosin, light polypeptide 4, alkali, atrial, embryonic	0.000811
57	1962_at	M14502	ARG1, arginase, liver; arginine degradation in the urea cycle	0.000811
58	31930_f_at	X63096	RHCE, Rhesus blood group, CcEe antigens	0.000978
59	34961_at	M88282	TACTILE, T cell activation, increased late expression	0.001179
60	36980_at	U03105	B4-2, prolin-rich protein with nuclear targeting signal	0.001541
61	38051_at	X76220	MAL, mal, T-cell differentiation protein	0.001546
62	1115_at	M25897	PF4, platelet factor 4	0.001546
63	33516_at	V00505	HBD, hemoglobin, delta	0.001551
64	32786_at	X51345	JUNB, jun B proto-oncogene; sequence specific transcription factor, BZIP family, JUN subfamily	0.001551
65	39209_r_at	M54995	PPBP, pro-platelet basic protein (includes platelet basic protein, beta-thromboglobulin, and others)	0.001747

66	677_s_at	J04430	ACP5, acid phosphatase 5, tartrate resistant	0.001747
67	33336_at	M27819	SLC4A1, solute carrier family 4, anion exchanger, member 1 (erythrocyte membrane protein band 3)	0.001935
68	37054_at	J04739	BPI, bactericidal /permeability-increasing protein	0.001941
69	32035_at	M16942	HLA-DRB4, major histocompatibility complex, class II, DR beta 4	0.001941
70	1992_at	U46922	FHIT, fragile histidine triad gene; nucleotide metabolism, tumor suppressor	0.001941
71	1971_g_at	U46922	FHIT, fragile histidine triad gene; nucleotide metabolism, tumor suppressor	0.001941
72	1223_at	X66362	PCTK3, PCTAIRE protein kinase 3	0.002413
73	37026_at	AF001461	COPEB, core promoter element binding protein; Krueppel family of zinc-finger proteins	0.00242
74	33121_g_at	AF045229	RGS10, regulator of G-protein signalling 10; signal transduction	0.00242
75	1894_f_at	L27065	NF2, neurofibromatosis 2 tumor suppressor	0.00242

* AMLs with partial duplication were not included in the analysis.

† Several genes are blood related and are probably due to contamination of some of the samples with blood

Finding Genes Common to Both ALLs and AMLs with MLL Chromosome Translocations

Having found genes whose expression pattern was different between MLL and ALLs with no translocations and between AML with MLL translocation and AMLs with no translocation, we intersected the results of the two tests (we used FDR 15% for each of the tests), in order to find genes in common. We identified 52 such genes (Table 3). The overexpressed genes included the phosphatase PPP2R5C, and the MCM4 gene whose product is an essential component of the prereplicative complex (You et al., 2002). The underexpressed genes included FHIT and JUNB.

Genes obtained from such an intersection are probably involved in common pathways in which various translocations of MLL lead to cancer.

Table 3 | Genes with similar behavior in ALLs and AMLs with MLL rearrangement*. At FDR=0.15

OVEREXPRESSED in tumors with MLL rearrangements

Gene ID	Acc. No	Symbol	Title and Function	p-value
176_at	U37352	PPP2R5C	protein phosphatase 2, regulatory subunit B (B56), gamma isoform	0.000002
37663_at	X70649	DDX1	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 1; DEAD box helicase family	0.000015
35255_at	AF098799	RANBP7	RAN binding protein 7	0.000015
37985_at	L37747	LMNB1	lamin B1; component of the nuclear lamina	0.000024
131_at	X83928	TAF21	TAF21, TATA box binding protein (TBP)-associated factor, 28 kD; TFIID complex	0.000028
35318_at	AB007944	KIAA0475	KIAA0475 gene product	0.000029
40701_at	U75362	USP13	ubiquitin specific protease 13 (isopeptidase T-3)	0.000031
981_at	X74794	MCM4	MCM4 minichromosome maintenance deficient 4 (S. cerevisiae); DNA replication	0.000049
40090_at	AI797997	WBSCR20	Williams Beuren syndrome chromosome region 20	0.000049
34936_at	AB012130	SLC4A7	solute carrier family 4, sodium bicarbonate cotransporter, member 7	0.000061

40786_at	U37352	PPP2R5C	protein phosphatase 2, regulatory subunit B (B56), gamma isoform	0.000062
38984_at	AB007896	KIAA0436	putative L-type neutral amino acid transporter	0.000080
335_r_at	L21990	SF3A2	splicing factor 3a, subunit 2, 66 kD (SAP62); binding of SNRNP complex to the BPS in mRNA	0.000116
33247_at	U86782	POH1	26S proteasome-associated pad1 homolog	0.000130
379_at	AB006679	APACD	ATP binding protein associated with cell differentiation	0.000135
40086_at	D87450	KIAA0261	KIAA0261 protein	0.000141
36968_s_at	AL050353	OIP2	Opa-interacting protein 2	0.000147
41480_at	AF029669	RAD51C	RAD51 homolog C (S. cerevisiae); DNA repair	0.000148
40459_at	S69189	ACOX1	acyl-Coenzyme A oxidase 1, palmitoyl; energy pathways, lipid metabolism	0.000153
2012_s_at	U34994	PRKDC	protein kinase, DNA-activated, catalytic polypeptide; DNA repair, recombination	0.000166
37304_at	U35451	CBX1	chromobox homolog 1 (HP1 beta homolog Drosophila); chromatin remodeling	0.000181
1798_at	U41060	LIV-1	LIV-1 protein, estrogen regulated	0.000186
38687_at	AL050051	DKFZP566D193	DKFZP566D193 protein	0.000194
41557_at	D29641	KIAA0052	KIAA0052 protein	0.000246
842_at	U48251	PRKCBP1	protein kinase C binding protein 1	0.000254
34768_at	AL080080	TXNDC	thioredoxin domain containing	0.000266
36099_at	M69040	SFRS1	splicing factor, arginine/serine-rich 1 (splicing factor 2, alternate splicing factor)	0.000276
34737_at	AF058718	COG5	component of oligomeric golgi complex 5	0.000279
39328_at	M11058	HMGCR	3-hydroxy-3-methylglutaryl-Coenzyme A reductase; sterol synthesis	0.000329
1448_at	D00762	PSMA3	proteasome (prosome, macropain) subunit, alpha type, 3; proteinase complex	0.000336
39788_at	X81889	PKP4	plakophilin 4; junctional plaques	0.000348
1850_at	U07418	MLH1	mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli); BRCA1-associated complex	0.000391
39432_at	AF038662	B4GALT4	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 4; glycosylation	0.000421
37359_at	D14658	KIAA0102	KIAA0102 gene product	0.000455
38349_at	AF038564	ITCH	itchy homolog E3 ubiquitin protein ligase (mouse)	0.000611
32777_at	Y12478	WRB	tryptophan rich basic protein	0.000613
35223_at	AB023234	AIBP63	alpha integrin binding protein 63	0.000674
38711_at	AB014527	CLASP2	cytoplasmic linker associated protein 2	0.000792
1515_at	L37374	FEN1	flap structure-specific endonuclease 1; cleaves 5' FLAP DNA structures, XPG/RAD2 family	0.001349
36478_at	X83973	TTF1	transcription termination factor, RNA polymerase I	0.002113
UNDEREXPRESSED in tumors with MLL rearrangements				
1992_at	U46922	FHIT	fragile histidine triad; nucleotide metabolism, tumor suppressor	0.000000
1971_g_at	U46922	FHIT	fragile histidine triad; nucleotide metabolism, tumor suppressor	0.000002
34904_at	S40369	GRIK5	Glutamate receptor, ionotropic, kainate 5; neurotransmission	0.000005
32786_at	X51345	JUNB	jun B proto-oncogene; sequence specific transcription factor	0.000010
34210_at	N90866	CDW52	CDW52 antigen (CAMPATH-1 antigen)	0.000033
1105_s_at	M12886	TRB	T cell receptor, beta	0.000073
33261_at	M16941	HLA-DRB1	major histocompatibility complex, class II, DR beta 1	0.000120
1007_s_at	U48705	DDR1	discoidin domain receptor family, member 1; tyrosine kinase, insulin receptor subfamily	0.000201
529_at	U15932	DUSP5	dual specificity phosphatase 5	0.000437
34491_at	AJ225089	OASL	2'-5'-oligoadenylate synthetase-like; binds RNA and DNA	0.000708
AFFX-M27830	M27830	28S rRNA	28S ribosomal RNA gene	0.001901
266_s_at	L33930	CD24	CD24 antigen (small cell lung carcinoma cluster 4 antigen); B-cell response	0.003808

* AMLs with MLL partial duplication were not included in the analysis.

Applying Two way Anova to Disease and Translocation data

After finding genes which behave similarly in both AMLs and ALLs with MLL chromosomal translocations, we turned to identify genes which are specific to only one type of disease and translocation state. Two way ANOVA was used on the samples, with the 2 dimensions being the disease type (ALL or AML), and the translocation status (presence or absence of the MLL chromosome translocation). Thus, each sample can belong to one of four groups:

	ALL	**AML**
No translocation	ALL with no Translocation	AML with no Translocation
translocation	ALL with Translocation	AML with Translocation

The genes that we would expect to retrieve are genes that are behaving inconsistently between AML and ALL, with the same rearrangement status. This inconsistency is a result of an **interaction** between the two factors *disease* and *Translocation status*.

The null hypothesis of the two-way ANOVA is that there is no interaction between the disease type and the translocation status with respect to their effect on the expression level of the tested gene. In the case of no interaction, the expected expression values of the gene in each of the four groups is the additive combination of the contributions of the effect of the disease type and of the effect of the rearrangement status. If the true expression values are far from the expected values, then we say that there is an interaction between the two factors regarding that gene's expression level.

Two possible mechanisms of interaction are an AND mechanism and an OR mechanism. In an AND mechanism the gene will be transcribed (or repressed) only in the presence of two different regulatory factors, which together mediate its transcription (or repression). In an OR mechanism the gene's transcription (or repression) can be either due to the presence of one factor or due to the presence of another factor. Therefore, by comparison to

the expression level in the *Normal* samples we may be able to say something about these genes mechanism of regulation.

Using FDR threshold of 5%, we found 49 genes whose expression levels are affected from an interaction between the disease type and rearrangement status. At the second stage of the analysis, we used unsupervised clustering of the genes in order to group them into clusters of similarly behaving genes. The genes were clustered according to their means in each of the four groups together with their mean in the normal samples. The mean expression levels of the ordered 49 genes are displayed in figure 16. The full list of genes is in the Appendix A table 7. Note that AMLs with MLL partial duplications were not included in the analysis.

The genes clustered into six distinct clusters, with six distinct behaviors (Figure 11, lists of genes are in Appendix A, tables 8-13):

(1) 4 Genes are upregulated only in AML with translocation. (Figures 10a, 11)

(2) 9 Genes are upregulated only in ALL without translocations (Figures 10b, 11).

(3) 5 Genes are upregulated only in ALL with translocations (Figures 10c, 11).

(4) 9 Genes are downregulated in both AMLs and in the ALL with translocation (Figures 10d, 11).

(5) 6 Genes are upregulated in both AMLs and in the ALL with translocation (Figures 10e, 11).

(6) 5 Genes which are downregulated in ALL without a translocation, upregulated in ALL with a translocation, and slightly upregulated in AML with no translocations (Figures 10f, 11).

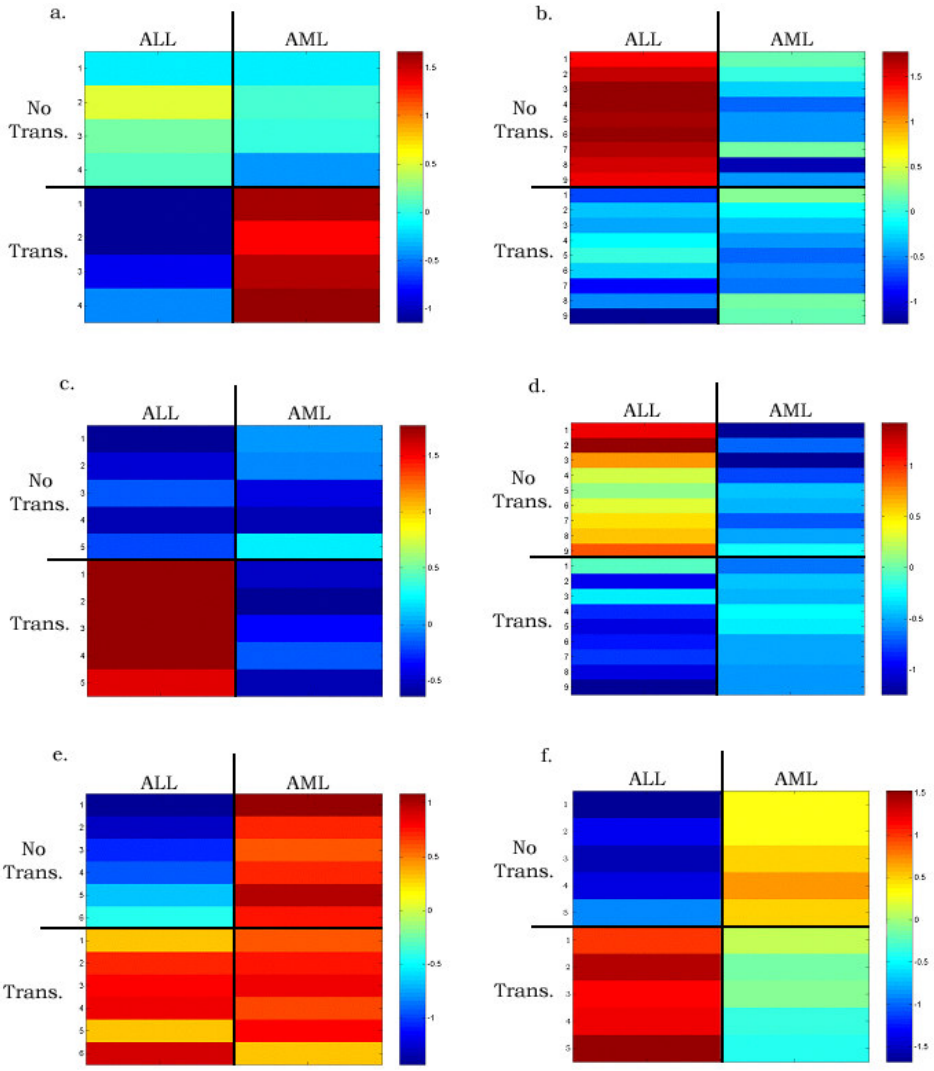

Figure 10 | Clusters of genes. **a**| cluster 1. Genes which are upregulated in AML with MLL translocation. **b**| cluster 2. Genes which are upregulated in ALL with no MLL translocation. **c**| cluster 3. Genes which are upregulated in ALLs with MLL translocation. **d**| cluster 4. Genes which are downregulated in AMLs and ALL with MLL translocation.. **e**| cluster 5. Genes which are upregulated in both AMLs and in ALL with MLL translocation. **f**| cluster 6. Genes which are downregulated in ALL with no MLL translocation, upregulated in ALL with MLL translocation and slightly upregulated in AML with no translocation.

Note that unless we know the expression levels of the genes in the *Normal* group, we cannot differentiate between the OR and the AND mechanisms. The mean expression levels of the 49 ordered genes, in each of the four groups together with their mean in the normal samples is shown in figure 16, the different clusters are indicated.

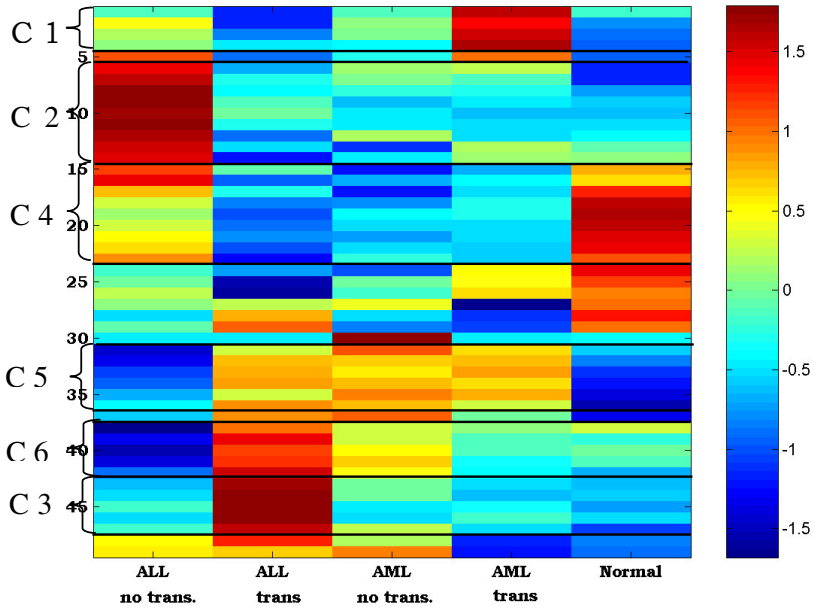

Figure 11 | The Mean Expression Levels of the Ordered Genes Found by the Two-Way ANOVA. Each row is centered and normalized. The clusters mentioned in the text are marked C1-C6

An example of an AND mechanism could be cluster 2: these genes need the presence of two factors in order to be expressed; one is present in samples of the ALL disease and the other is found where there is no MLL translocation. The two factors are only expressed together in ALL without MLL translocation, and therefore these genes are upregulated only in this group of samples.

An example of an OR mechanism might be cluster 4: The genes of cluster 4 are regulated by two different repressors, each one acting alone. One of these repressors is expressed in AMLs and the other is expressed in leukemias with MLL translocations. Therefore, the genes are downregulated in AMLs and ALL with translocations compared to *Normal* samples and ALL without translocations.

An anticipated direction for this work would be to check for shared regulatory elements between genes of the same cluster, in order to find common transcription regulators, which can be later connected to a certain model depending on their expression in the different conditions.

Unsupervised Analysis

For the unsupervised analysis, we utilized the CTWC method (see methods section). Among the clusters found using this method, there were a few clusters of genes, separating the samples in an interesting way. We will focus on two of those.

Cluster 7: Genes Partitioning the ALL samples into t(4:11) and the rest of the ALL samples.

This cluster of genes was found after clustering the set of all the genes, denoted G1, using samples of cluster S5; this clustering process is denoted G1(S5) (Figure 13). S5 was found when all the samples, S1, were clustered using all the genes, G1, this process is therefore denoted S1(G1) (Figure 12). S5 contained 11 t(4:11) ALLs and two CD10- ALL samples.

Clustering of the ALL samples with G7 resulted in a very stable cluster of 13 t(4:11) and two CD10- ALLs (Figure 14).

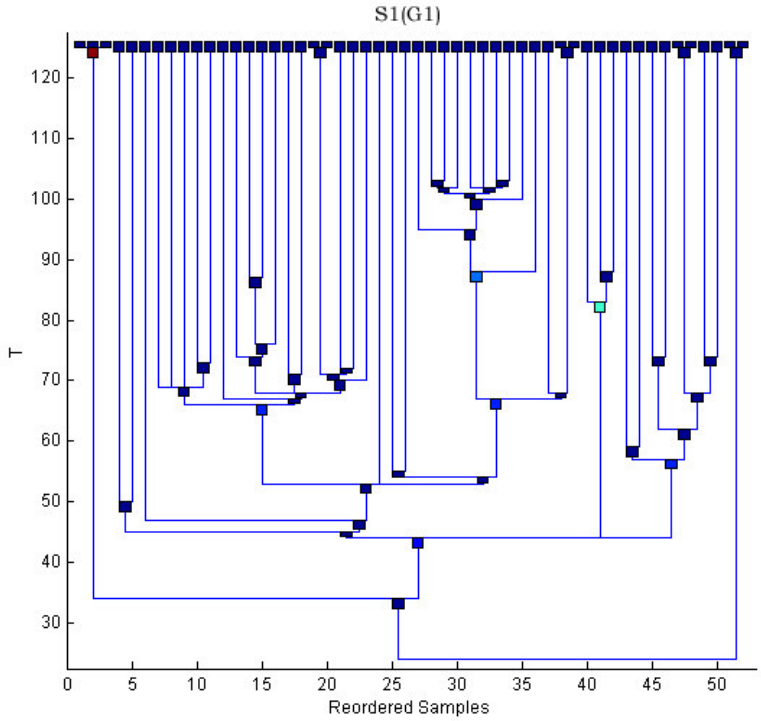

Figure 12 | **Dendrogram of S1(G1)**. This clustering procedure yielded the sample cluster S5 marked in pink.

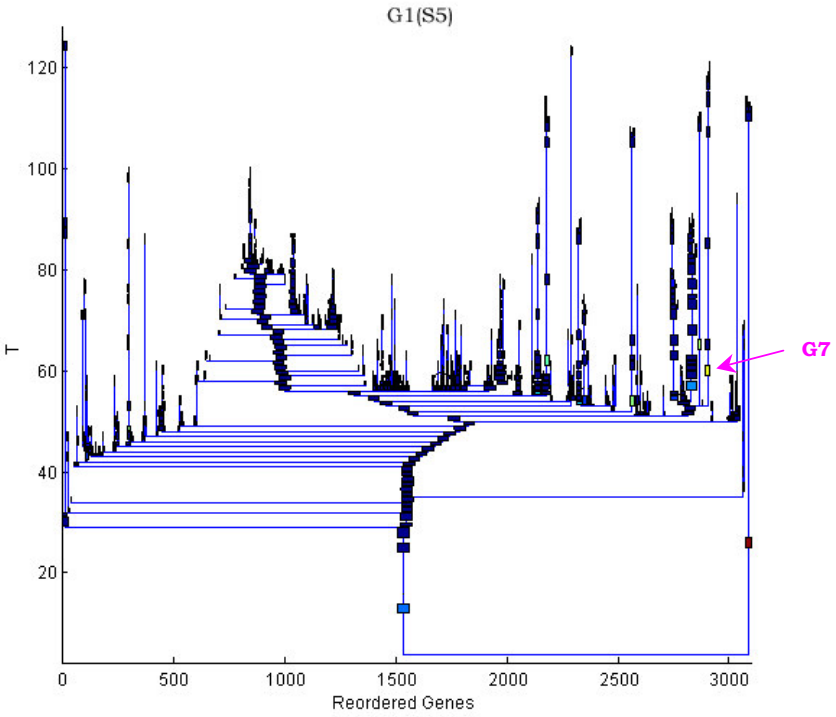

Figure 13 | Dendrogram of G1(S5). This clustering procedure yielded the gene cluster G7 marked in pink.

These genes were found to be very low on the t(4:11) and two out of the 4 CD10-, but having an inconsistent behavior on the rest of the ALL samples (see table 14 in Appendix A for the list of genes).

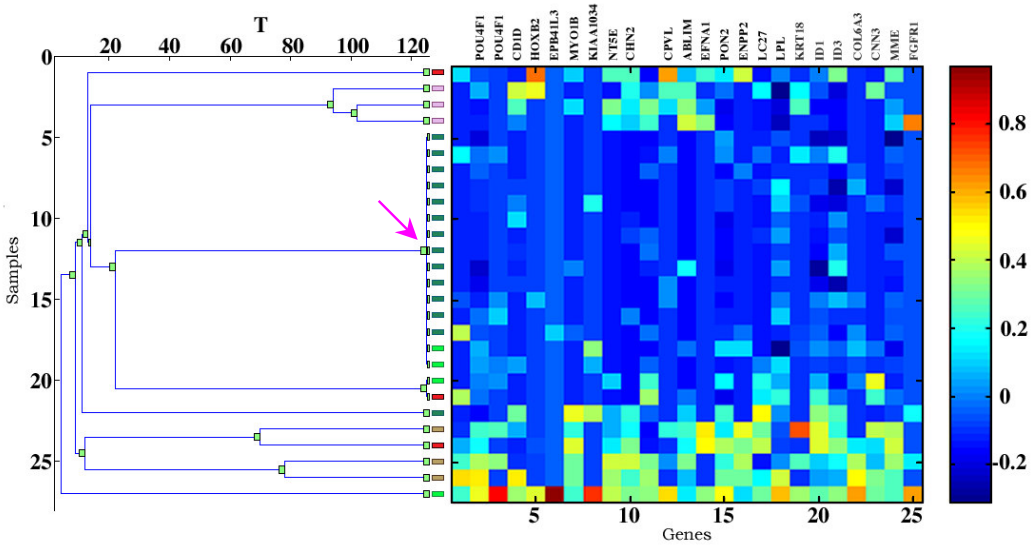

Fig 14 | Clustering of the ALL Samples According to the Genes of Cluster 7. Rows are samples, columns are genes. The t(4:11) ALL and the CD10- are in dark and light green respectively. The t(4:11) cluster is indicted by a pink arrow.

When comparing these genes to the results of the supervised analysis comparing t(4:11) ALLs to ALLs with no translocations (at FDR level 5%), only 5 of these genes were found (POU4F1, NT5E, ABLIM, CNN3, MME). These genes were not discovered by the supervised test since their expression levels were not consistent in the non-t(4:11) samples, being dispersed and therefore not rendering a small enough p-value in the Ranksum test. On the other hand, the CTWC was able to find these genes because of their high correlation in some of the samples. The possible biological meaning of these genes is that while in the non-t(4:11) samples the values are randomly dispersed, in the t(4:11) samples the genes are consistently downregulated. This behavior maybe a general phenomenon relevant to biological systems: while in some population of samples the genes' expressions maybe fluctuating because of time-dependency or the inhomogenousity of the samples, in the other population the genes maybe strongly regulated.

Cluster 8: Genes Separating the Samples into Cell-line samples and Primary Tumors.

Using the CTWC method a cluster of 132 genes was found, separating the cell-lines from the primary tumors. This cluster of genes was found after clustering all the genes, denoted G1, using the set of all samples, denoted S1, this clustering process is denoted G1(S1) (Figure 15).

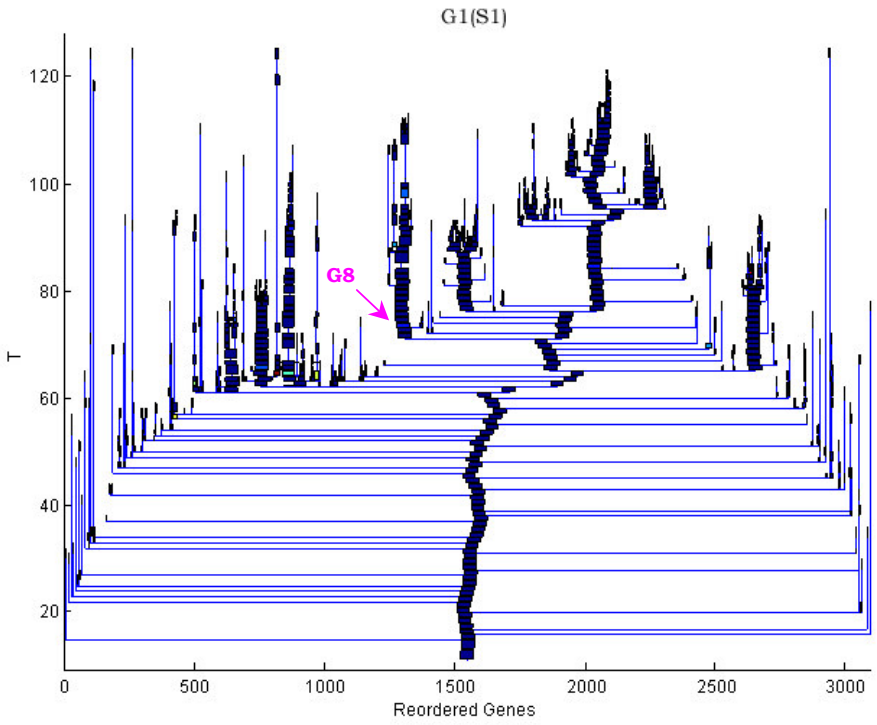

Figure 15 | **Dendrogram of G1(S1)**. This clustering procedure yielded the gene cluster G8 marked in pink.

The partitioning of the samples using this gene cluster was very stable, with the cell-lines diverging from the rest of the samples very clearly. This separation was not anticipated by us, only upon further investigations we learned that these 7 samples are cell lines. The expression level of these genes does not depend on the cancer type, these genes are expressed at low levels in all the cell-lines (AMLs and ALLs) and at higher levels in all primary tumors. Many of these genes were genes coding for histones and there were also various genes related to immunoglobulins. The complete list of genes can be found at the Appendix A table 15.

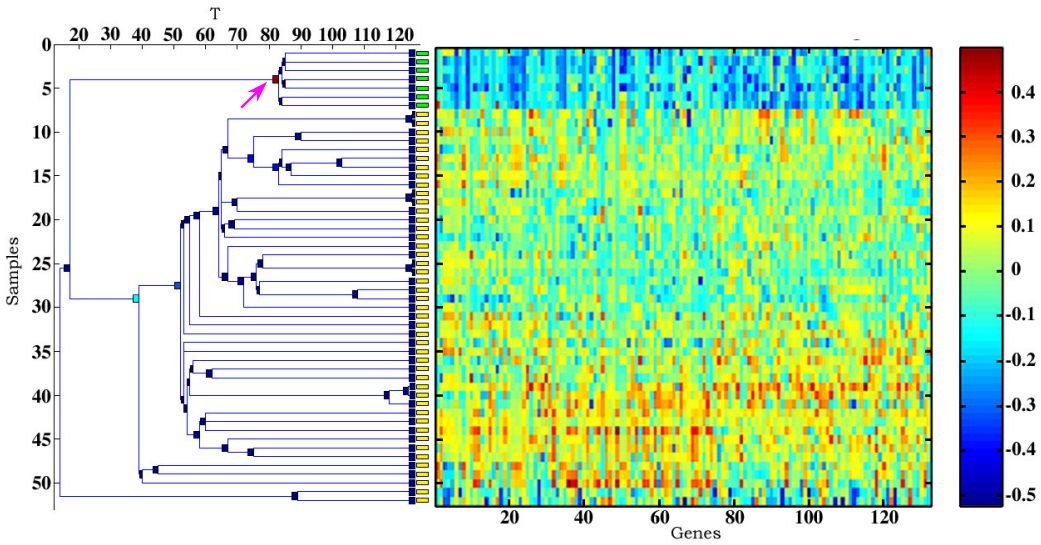

Figure 16 | Clustering of the Samples According to the Genes of Cluster 8. Rows are samples, columns are genes. The cell-lines are in green. The cell-line cluster is indicated by a pink arrow.

Thus, CTWC is a powerful method enabling us to discover unexpected partitions in the data. Furthermore, CTWC was also able to find genes that separated the data in a known way, most of these genes were not found in an equivalent supervised analysis at FDR level of 5%.

From this we can conclude that the two methods are complementary to each other, and both should be used when analyzing gene expression data.

Reference List

Alberts,B., Bray,D., Lewis,J., Raff,M., Roberts,K., and Watson,J.D. (1994). Molecular Biology of the Cell. (New York and London: Garland Publishing).

Armstrong,S.A., Staunton,J.E., Silverman,L.B., Pieters,R., den Boer,M.L., Minden,M.D., Sallan,S.E., Lander,E.S., Golub,T.R., and Korsmeyer,S.J. (2002). MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat. Genet. *30*, 41-47.

Ayton,P.M. and Cleary,M.L. (2001). Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins. Oncogene *20*, 5695-707.

Bast,R.C., Kufe,D.W., Pollock,R.E., Weichselbaum,R.R., Holland,J.F., and Emil,F. (2002). Neoplasms of the Hematopoitic System, Acute Myeloid Leukemia in Adults. In Cancer Medicine, BC Decker Inc).

Berger,R., Chen,S.J., and Chen,Z. (1990). Philadelphia-positive acute leukemia. Cytogenetic and molecular aspects. Cancer Genet. Cytogenet. *44*, 143-152.

Bernstein,B.E., Humphrey,E.L., Erlich,R.L., Schneider,R., Bouman,P., Liu,J.S., Kouzarides,T., and Schreiber,S.L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. Proc. Natl. Acad. Sci. U. S. A *99*, 8695-8700.

Blackshaw,S., Sawa,A., Sharp,A.H., Ross,C.A., Snyder,S.H., and Khan,A.A. (2000). Type 3 inositol 1,4,5-trisphosphate receptor modulates cell death. FASEB J. *14*, 1375-1379.

Breen,T.R. and Harte,P.J. (1993). Trithorax regulates multiple homeotic genes in the bithorax and Antennapedia complexes and exerts different tissue-specific, parasegment-specific and promoter-specific effects on each. Development *117*, 119-34.

Brown,B.A. (1993). **Hematology: Principles and Procedures**. (Philadelphia: Lea & Febiger).

Burmeister,T. and Thiel,E. (2001). Molecular genetics in acute and chronic leukemias. J. Cancer Res. Clin. Oncol. *127*, 80-90.

Cairns,B.R., Henry,N.L., and Kornberg,R.D. (1996). TFG/TAF30/ANC1, a component of the yeast SWI/SNF complex that is similar to the leukemogenic proteins ENL and AF-9. Mol. Cell Biol. *16*, 3308-3316.

Canaani,E. and Gale,R.P. (1985). Association between the Philadelphia chromosome and a unique abl transcript. Hamatol Bluttransfus *29*, 154-8.

Canaani,E., Gale,R.P., Steiner-Saltz,D., Berrebi,A., Aghai,E., and Januszewicz,E. (1984). Altered transcription of an oncogene in chronic myeloid leukaemia. Lancet *1*, 593-5.

Cimino,G., Moir,D.T., Canaani,O., Williams,K., Crist,W.M., Katzav,S., Cannizzaro,L., Lange,B., Nowell,P.C., Croce,C.M., and . (1991). Cloning of ALL-1, the locus involved in leukemias with the t(4;11)(q21;q23), t(9;11)(p22;q23), and t(11;19)(q23;p13) chromosome translocations. Cancer Res. *51*, 6712-6714.

Cleary,M.L. (1991). Oncogenic conversion of transcription factors by chromosomal translocations. Cell *66*, 619-622.

Corral,J., Lavenir,I., Impey,H., Warren,A.J., Forster,A., Larson,T.A., Bell,S., McKenzie,A.N., King,G., and Rabbitts,T.H. (1996). An Mll-AF9 fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes. Cell *85*, 853-861.

Daley,G.Q., Van Etten,R.A., and Baltimore,D. (1990). Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. Science *247*, 824-830.

Dobson,C.L., Warren,A.J., Pannell,R., Forster,A., Lavenir,I., Corral,J., Smith,A.J., and Rabbitts,T.H. (1999). The mll-AF9 gene fusion in mice controls myeloproliferation and specifies acute myeloid leukaemogenesis. EMBO J. *18*, 3564-3574.

Dobson,C.L., Warren,A.J., Pannell,R., Forster,A., and Rabbitts,T.H. (2000). Tumorigenesis in mice with a fusion of the leukaemia oncogene Mll and the bacterial lacZ gene. EMBO J. *19*, 843-851.

Druker,B.J., Sawyers,C.L., Kantarjian,H., Resta,D.J., Reese,S.F., Ford,J.M., Capdeville,R., and Talpaz,M. (2001). Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. N. Engl. J. Med. *344*, 1038-1042.

Eistere,W., Hilbe,W., Stauder,R., Bechter,O., Fend,F., and Thaler,J. (1996). An aggressive subtype of B-CLL is characterized by strong CD44 expression and lack of CD11c. Br. J. Haematol. *93*, 661-669.

Fainstein,E., Marcelle,C., Rosner,A., Canaani,E., Gale,R.P., Dreazen,O., Smith,S.D., and Croce,C.M. (1987). A new fused transcript in Philadelphia chromosome positive acute lymphocytic leukaemia. Nature *330*, 386-388.

Fujita,Y., Kitagawa,M., Nakamura,S., Azuma,K., Ishii,G., Higashi,M., Kishi,H., Hiwasa,T., Koda,K., Nakajima,N., and Harigaya,K. (2002). CD44 signaling through focal adhesion kinase and its anti-apoptotic effect. FEBS Lett. *528*, 101-108.

Gale,R.P. and Canaani,E. (1984). An 8-kilobase abl RNA transcript in chronic myelogenous leukemia. Proc Natl Acad Sci U S A *81*, 5648-52.

Galoian,K., Milne,T., Brock,H.W., Shilatifard,A., Hess,J.L., and Slany,R. (2000). Deregulation of c-MYC by leukemogenic MLL fusion proteins. Blood *96*, 457a.

Gil,J. and Esteban,M. (2000). Induction of apoptosis by the dsRNA-dependent protein kinase (PKR): mechanism of action. Apoptosis. *5*, 107-114.

Gilliland,D.G. and Tallman,M.S. (2002). Focus on acute leukemias. Cancer Cell *1*, 417-20.

Groffen,J., Stephenson,J.R., Heisterkamp,N., de Klein,A., Bartram,C.R., and Grosveld,G. (1984). Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. Cell *36*, 93-99.

Harvey,L., Arnold,B., Lawrence,S.Z., Paul,M., David,B., and James,D. (2002). Cancer, Oncogenic Mutations Affecting Cell Proliferation. In Molecular Cell Biology, (New York: W. H. FREEMAN).

Hata,A., Shi,Y., and Massague,J. (1998). TGF-beta signaling and cancer: structural and functional consequences of mutations in Smads. Mol. Med. Today *4*, 257-262.

Heerema,N.A., Arthur,D.C., Sather,H., Albo,V., Feusner,J., Lange,B.J., Steinherz,P.G., Zeltzer,P., Hammond,D., and Reaman,G.H. (1994). Cytogenetic features of infants less than 12 months of age at diagnosis of acute lymphoblastic leukemia: impact of the 11q23 breakpoint on outcome: a report of the Childrens Cancer Group. Blood *83*, 2274-84.

Heisterkamp,N., Stephenson,J.R., Groffen,J., Hansen,P.F., de Klein,A., Bartram,C.R., and Grosveld,G. (1983). Localization of the c-ab1 oncogene adjacent to a translocation break point in chronic myelocytic leukaemia. Nature *306*, 239-242.

Huret, J. L. AF10 (ALL1 fused gene from chromosome 10). Atlas Genet Cytogenet Oncol Haematol. 1997a.
Ref Type: Electronic Citation

Huret, J. L. AF6 (ALL1 fused gene from chromosome 6). Atlas Genet Cytogenet Oncol Haematol. 1997b.
Ref Type: Electronic Citation

Huret, J. L. AF9 (ALL1 fused gene from chromosome 9). Atlas Genet Cytogenet Oncol Haematol. 1997c.
Ref Type: Electronic Citation

Huret, J. L. ENL (eleven nineteen leukemia). Atlas Genet Cytogenet Oncol Haematol. 1997d.

Ref Type: Electronic Citation

Huret, J. L. and Marschalek, R. AF4 (ALL1 fused gene from chromosome 4). Atlas Genet Cytogenet Oncol Haematol. 2002.
Ref Type: Electronic Citation

Isnard,P., Core,N., Naquet,P., and Djabali,M. (2000). Altered lymphoid development in mice deficient for the mAF4 proto-oncogene. Blood *96*, 705-710.

Jang,C.W., Chen,C.H., Chen,C.C., Chen,J.Y., Su,Y.H., and Chen,R.H. (2002). TGF-beta induces apoptosis through Smad-mediated expression of DAP-kinase. Nat. Cell Biol. *4*, 51-58.

Kaneko,Y., Maseki,N., Takasaki,N., Sakurai,M., Hayashi,Y., Nakazawa,S., Mori,T., Takeda,T., Shikano,T., and et al. (1986). Clinical and hematologic characteristics in acute leukemia with 11q23 translocations. Blood *67*, 484-91.

Kroon,E., Krosl,J., Thorsteinsdottir,U., Baban,S., Buchberg,A.M., and Sauvageau,G. (1998). Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. EMBO J. *17*, 3714-3725.

Lachner,M., O'Carroll,D., Rea,S., Mechtler,K., and Jenuwein,T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Lavau,C., Szilvassy,S.J., Slany,R., and Cleary,M.L. (1997). Immortalization and leukemic transformation of a myelomonocytic precursor by retrovirally transduced HRX-ENL. EMBO J. *16*, 4226-4237.

Leukemia-Lymphoma Society. **Leukemia**. The Leukemia Lymphoma Society . 2002.
Ref Type: Electronic Citation

Li,X., Scuderi,A., Letsou,A., and Virshup,D.M. (2002). B56-associated protein phosphatase 2A is required for survival and protects from apoptosis in Drosophila melanogaster. Mol. Cell Biol. *22*, 3674-3684.

Liebermann,D.A., Gregory,B., and Hoffman,B. (1998). AP-1 (Fos/Jun) transcription factors in hematopoietic differentiation and apoptosis. Int. J. Oncol *12*, 685-700.

Liesner,R.J. and Goldstone,A.H. (1997). ABC of clinical haematology. The acute leukaemias. Bmj *314*, 733-6.

Look,A.T. (1997). Oncogenic transcription factors in the human acute leukemias. Science *278*, 1059-64.

Maacke,H., Opitz,S., Jost,K., Hamdorf,W., Henning,W., Kruger,S., Feller,A.C., Lopens,A., Diedrich,K., Schwinger,E., and Sturzbecher,H.W.

(2000). Over-expression of wild-type Rad51 correlates with histological grading of invasive ductal breast cancer. Int. J. Cancer *88*, 907-913.

Mecucci,C., Rosati,R., and Starza,R.L. (2002). Genetic profile of acute myeloid leukemia. Rev. Clin. Exp. Hematol. *6*, 3-25.

Milne,T.A., Briggs,S.D., Brock,H.W., Martin,M.E., Gibbs,D., Allis,C.D., and Hess,J.L. (2002). MLL Targets SET Domain Methyltransferase Activity to Hox Gene Promoters. Mol. Cell *10*, 1107-1117.

Nagy,P.L., Griesenbeck,J., Kornberg,R.D., and Cleary,M.L. (2002). A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. Proc. Natl. Acad. Sci. U. S. A *99*, 90-94.

Nakamura,T., Alder,H., Gu,Y., Prasad,R., Canaani,O., Kamada,N., Gale,R.P., Lange,B., Crist,W.M., Nowell,P.C., and et al. (1993). Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leukemia share sequence homology and/or common motifs. Proc Natl Acad Sci U S A *90*, 4631-5.

Nakamura,T., Largaespada,D.A., Shaughnessy,J.D., Jr., Jenkins,N.A., and Copeland,N.G. (1996). Cooperative activation of Hoxa and Pbx1-related genes in murine myeloid leukaemias. Nat. Genet *12*, 149-153.

Nakamura,T., Mori,T., Tada,S., Krajewski,W., Rozovskaia,T., Wassell,R., Dubois,G., Mazo,A., Croce,C.M., and Canaani,E. (2002). ALL-1 Is a Histone Methyltransferase that Assembles a Supercomplex of Proteins Involved in Transcriptional Regulation. Mol Cell *10*, 1119-28.

Nesbit,C.E., Tersak,J.M., and Prochownik,E.V. (1999). MYC oncogenes and human neoplastic disease. Oncogene *18*, 3004-3016.

Noma,K., Allis,C.D., and Grewal,S.I. (2001). Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science *293*, 1150-1155.

Noma,K. and Grewal,S.I. (2002). Histone H3 lysine 4 methylation is mediated by Set1 and promotes maintenance of active chromatin states in fission yeast. Proc. Natl. Acad. Sci. U. S. A *99 Suppl 4*, 16438-16445.

Pallardy,M. and Biola,A. (1998). [Induction of apoptosis in lymphocytes by glucocorticoids: between physiology and pharmacology]. C. R. Seances Soc. Biol. Fil. *192*, 1051-1063.

Passegue,E., Jochum,W., Schorpp-Kistner,M., Mohle-Steinlein,U., and Wagner,E.F. (2001). Chronic myeloid leukemia with increased granulocyte progenitors in mice lacking junB expression in the myeloid lineage. Cell *104*, 21-32.

Pedersen-Bjergaard,J. and Rowley,J.D. (1994). The balanced and the unbalanced chromosome aberrations of acute myeloid leukemia may

develop in different ways and may contribute differently to malignant transformation. Blood *83*, 2780-2786.

Pekarsky,Y., Zanesi,N., Palamarchuk,A., Huebner,K., and Croce,C.M. (2002). FHIT: from gene discovery to cancer treatment and prevention. Lancet Oncol *3*, 748-754.

Rabinovich,G.A., Baum,L.G., Tinari,N., Paganelli,R., Natoli,C., Liu,F.T., and Iacobelli,S. (2002). Galectins and their ligands: amplifiers, silencers or tuners of the inflammatory response? Trends Immunol. *23*, 313-320.

Raveh,T., Droguett,G., Horwitz,M.S., DePinho,R.A., and Kimchi,A. (2001). DAP kinase activates a p19ARF/p53-mediated apoptotic checkpoint to suppress oncogenic transformation. Nat. Cell Biol. *3*, 1-7.

Reya,T., Morrison,S.J., Clarke,M.F., and Weissman,I.L. (2001). Stem cells, cancer, and cancer stem cells. Nature *414*, 105-11.

Rowley,J.D. (1973b). Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature *243*, 290-293.

Rowley,J.D. (1973a). Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature *243*, 290-293.

Rowley,J.D. (1998). The critical role of chromosome translocations in human leukemias. Annu Rev Genet *32*, 495-519.

Rowley,J.D. (2000). Molecular genetics in acute leukemia. Leukemia *14*, 513-517.

Rowley,J.D. (2001). Chromosome translocations: dangerous liaisons revisited. Nat Rev Cancer *1*, 245-50.

Rozenblatt-Rosen,O., Rozovskaia,T., Burakov,D., Sedkov,Y., Tillib,S., Blechman,J., Nakamura,T., Croce,C.M., Mazo,A., and Canaani,E. (1998). The C-terminal SET domains of ALL-1 and TRITHORAX interact with the INI1 and SNR1 proteins, components of the SWI/SNF complex. Proc Natl Acad Sci U S A *95*, 4152-7.

Rozovskaia,T., Feinstein,E., Mor,O., Foa,R., Blechman,J., Nakamura,T., Croce,C.M., Cimino,G., and Canaani,E. (2001). Upregulation of Meis1 and HoxA9 in acute lymphocytic leukemias with the t(4 : 11) abnormality. Oncogene *20*, 874-878.

Schichman,S.A., Caligiuri,M.A., Gu,Y., Strout,M.P., Canaani,E., Bloomfield,C.D., and Croce,C.M. (1994). ALL-1 partial duplication in acute leukemia. Proc. Natl. Acad. Sci. U. S. A *91*, 6236-6239.

Schichman,S.A., Canaani,E., and Croce,C.M. (1995). Self-fusion of the ALL1 gene. A new genetic mechanism for acute leukemia. Jama *273*, 571-6.

Shaulian,E. and Karin,M. (2001). AP-1 in cell proliferation and survival. Oncogene *20*, 2390-2400.

Shen,W.F., Rozenfeld,S., Kwong,A., Kom ves,L.G., Lawrence,H.J., and Largman,C. (1999). HOXA9 forms triple complexes with PBX2 and MEIS1 in myeloid cells. Mol. Cell Biol. *19*, 3051-3061.

Shtivelman,E., Lifshitz,B., Gale,R.P., and Canaani,E. (1985). Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. Nature *315*, 550-4.

Strahl,B.D. and Allis,C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Thorsteinsdottir,U., Sauvageau,G., Hough,M.R., Dragowska,W., Lansdorp,P.M., Lawrence,H.J., Largman,C., and Humphries,R.K. (1997). Overexpression of HOXA10 in murine hematopoietic cells perturbs both myeloid and lymphoid differentiation and leads to acute myeloid leukemia. Mol. Cell Biol. *17*, 495-505.

Tkachuk,D., Hirschmann,J.V., and McArthur,J.R. (2002). White Cell Disorders. In Atlas of Clinical Hematology, Elsevier Science).

Tkachuk,D.C., Kohler,S., and Cleary,M.L. (1992). Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias. Cell *71*, 691-700.

Tseng,Y.H., Ueki,K., Kriauciunas,K.M., and Kahn,C.R. (2002). Differential roles of insulin receptor substrates in the anti-apoptotic function of insulin-like growth factor-1 and insulin. J. Biol. Chem. *277*, 31601-31611.

Veldman,T., Vignon,C., Schrock,E., Rowley,J.D., and Ried,T. (1997). Hidden chromosome abnormalities in haematological malignancies detected by multicolour spectral karyotyping. Nat. Genet. *15*, 406-410.

Virginia University, Dept. of Pathology. Basic Hematology. University of Virginia . 2002.
Ref Type: Electronic Citation

Yamaoka,K., Mishima,K., Nagashima,Y., Asai,A., Sanai,Y., and Kirino,T. (2000). Expression of galectin-1 mRNA correlates with the malignant potential of human gliomas and expression of antisense galectin-1 inhibits the growth of 9 glioma cells. J. Neurosci. Res. *59*, 722-730.

Yano,T., Nakamura,T., Blechman,J., Sorio,C., Dang,C.V., Geiger,B., and Canaani,E. (1997). Nuclear punctate distribution of ALL-1 is conferred by distinct elements at the N terminus of the protein. Proc. Natl. Acad. Sci. U. S. A *94*, 7286-7291.

Yeoh,E.J., Ross,M.E., Shurtleff,S.A., Williams,W.K., Patel,D., Mahfouz,R., Behm,F.G., Raimondi,S.C., Relling,M.V., Patel,A., Cheng,C., Campana,D., Wilkins,D., Zhou,X., Li,J., Liu,H., Pui,C.H., Evans,W.E., Naeve,C., Wong,L., and Downing,J.R. (2002). Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell *1*, 133-143.

You,Z., Ishimi,Y., Masai,H., and Hanaoka,F. (2002). Roles of Mcm7 and Mcm4 subunits in the DNA helicase activity of the mouse Mcm4/6/7 complex. J. Biol. Chem. *277*, 42471-42479.

Yu,B.D., Hanson,R.D., Hess,J.L., Horning,S.E., and Korsmeyer,S.J. (1998). MLL, a mammalian trithorax-group gene, functions as a transcriptional maintenance factor in morphogenesis. Proc Natl Acad Sci U S A *95*, 10632-6.

Yu,B.D., Hess,J.L., Horning,S.E., Brown,G.A., and Korsmeyer,S.J. (1995). Altered Hox expression and segmental identity in Mll-mutant mice. Nature *378*, 505-8.

Zegerman,P., Canas,B., Pappin,D., and Kouzarides,T. (2002). Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. J. Biol. Chem. *277*, 11621-11624.

Zeleznik-Le,N.J., Harden,A.M., and Rowley,J.D. (1994). 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. Proc Natl Acad Sci U S A *91*, 10610-4.

)

Chapter 3

Gene Expression Analysis of Endochondral Bone Formation

Introduction

Bone Composition and Morphology

Bone is a highly specialized type of connective tissue in which the extracellular matrix is mineralized. This complex living tissue provides rigid support and protection for all higher vertebrates.

Bone is made of an organic matrix that is strengthened by deposits of calcium salts. The organic matrix is mostly type I collagen, which comprises about 95% of this matrix and the remaining 5% are composed of proteoglycans and various noncollagenous proteins.

Morphologically, bone can be either compact (cortical) or spongy (cancellous). Compact bone is densely packed, playing mechanical and protective roles, whereas cancellous bone has a loosely organized porous matrix and carries metabolic functions (Marks and Odgren, 2002).

The bone tissue is composed of four different cell types. Cells residing on the bone surface are osteoblasts, osteoclasts and bone lining cells, whereas osteocyte cells are found in the bone interior. Osteoblasts (reviewed by Ducy et al., 2000), osteocytes and bone lining cells are of *mesenchym* origins, while osteoclasts arise from *monocyte/ macrophage* origins by fusion of mononuclear progenitors (reviewed by Teitelbaum, 2000).

Osteoblasts are fully differentiated cells, in charge of the production of bone matrix. Osteoblasts secrete type I collagen and the various noncollagenic proteins of the bone matrix. Matrix deposition is usually polarized towards the surface of the bone, but once in a while, the secretion broadens, surrounding the osteoblast, turning it into an

osteocyte. Thus, the *osteocyte* is a mature osteoblast immersed within the bone matrix. The osteocyte cells are responsible for the maintenance of the matrix, having the capacity to resorb matrix (to some extent) as well as to produce it.

Bone Lining Cells are osteogenic progenitors, i.e., precursors to osteoblasts.

Osteoclasts are multinucleated cells responsible for bone resorbtion. The osteoclasts play a central role in the formation of the skeleton and regulation of its mass.

Bone Formation

During embryonic development, bone formation (ossification) occurs via two distinct pathways: (i) a direct process called *intramembranous ossification* or, (ii) an indirect process, *endochondral ossification* (Marks and Odgren, 2002).

Intramembranous ossification is a process that occurs during embryonic development to produce, amongst others, the flat bones of the skull. At first, the mesenchymal progenitor cells condense in specific areas prefiguring a future bone. Subsequently, the condensed area is invaded by blood vessels and the mesenchymal cells differentiate directly into osteoblasts, secreting bone matrix.

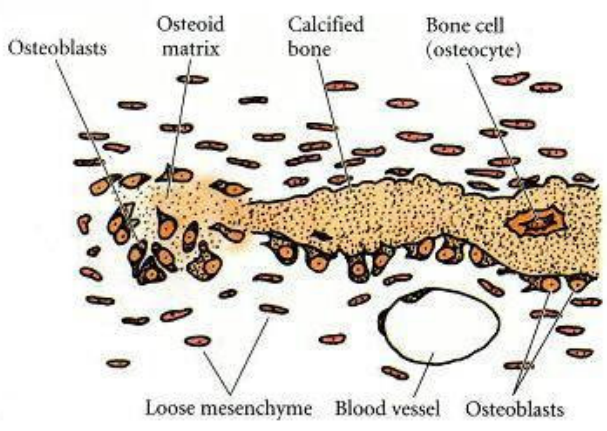

Figure 1 |Schematic diagram of intramembranous ossification. Mesenchymal cells condense to produce osteoblasts, which deposit osteoid matrix. These osteoblasts become arrayed along the calcified region of the matrix. Osteoblasts that are trapped within the bone matrix become osteocytes (adapted from Gilbert, 2000).

Endochondral ossification is the process by which weight-bearing bones and bones of the joints are created. In such a process mesenchymal stem cells migrate to the future bone site and condense as in the first stages of the direct ossification. However, the next steps of this process are different: instead of differentiating directly into bone, the mesenchym turns into a cartilage template, prefiguring the future bone (Horton, 1990). In the central part of the future bone align chondrocyte cells (cells producing cartilage), originating from mesenchymal cells, proliferate, undergo hypertrophy[1] and mineralize the extracellular matrix. This process is done in an orderly fashion, producing a cartilaginous scaffold for the future bone. Next, blood vessels penetrate into the surface (perichondrium) of the central part of the cartilage template, transforming the perichondrium cells into osteoblasts (either by introducing new osteoproginator cells from the blood or by differentiation of the chondroblasts to osteoblasts). These osteoblasts create a bone sleeve, over the cartilage template. Blood vessels further penetrate the scaffold interior, introducing osteoblasts to the cartilage core. The osteoblasts progressively replace the mineralized cartilage with bone matrix. The oseoblasts further differentiate into osteocytes. The chondrocytes disappear after the deposition of the cartilage, some differentiating into osteoblasts (Galotto et al., 1994; Roach et al., 1995; Thesingh et al., 1991) while others going through apoptosis.

This mechanism of bone development is also noticeable in bone growth and fracture repair, i.e., it is found in the adult as well as during development.

[1] Hypertrophy- An increase in the size of a tissue or organ that results from an increase in size of the cell present.

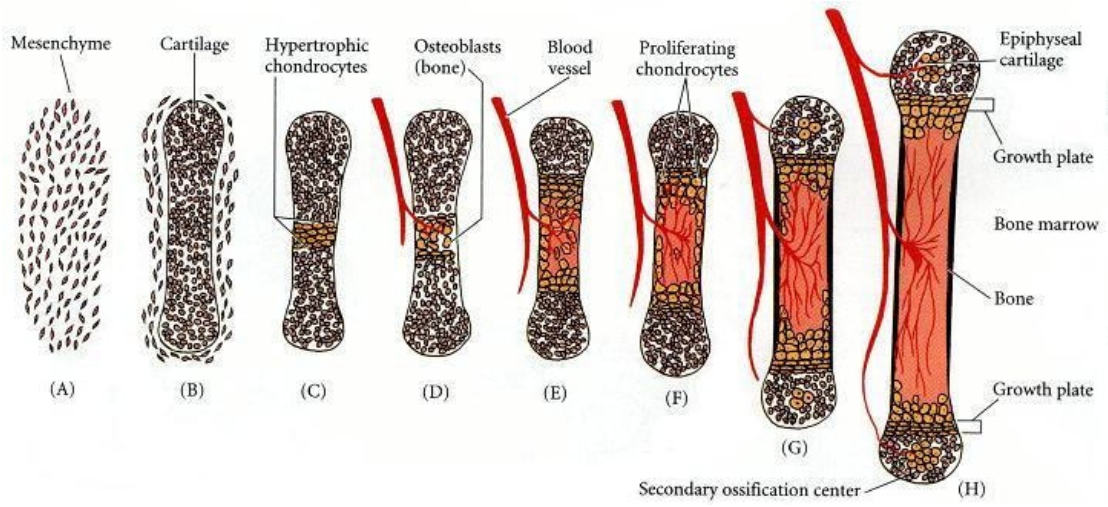

Figure 2 |Schematic diagram of endochondral ossification. (A, B) Mesenchymal cells condense and differentiate into chondrocytes to form the cartilaginous model of the bone. (C) Chondrocytes in the center of the shaft undergo hypertrophy and apoptosis while they change and mineralize their extracellular matrix. Their deaths allow blood vessels to enter. (D, E) Blood vessels bring in osteoblasts, which bind to the degenerating cartilaginous matrix and deposit bone matrix. (F-H) Bone formation and growth consist of ordered arrays of proliferating, hypertrophic, and mineralizing chondrocytes. Secondary ossification centers also form as blood vessels enter near the tips of the bone (adapted from Horton, 1990).

Mesenchymal Stem Cells

Mesenchymal stem cells (MSCs) are multipotent cells present in the bone marrow and in various adult mesenchymal tissues (fat, skin, muscle, etc.). MSCs have the capabilities to replicate while maintaining their current state as multipotent cells and also to differentiate into various tissue types, being responsible for the regeneration and repair of these tissues. MSCs have the potential to differentiate to lineages of mesenchymal tissues (Figure 3), including bone (Haynesworth et al., 1992; Pittenger et al., 1999; Liechty et al., 2000), cartilage (Pittenger et al., 1999; Yoo et al., 1998; Mackay et al., 1998; Liechty et al., 2000), tendon (Young et al., 1998), ligament, fat (Pittenger et al., 1999; Liechty et al., 2000) and a range of other connective tissues (Caplan, 1994). MSCs were shown to coexpress mRNAs specific to bone, cartilage, muscle, stroma, endothelia and neuron in the absence of differentiation-inducing agent (Tremain et al., 2001). In addition, MSC have also been shown to have some degree of plasticity being able to differentiate to glial cell of the

neuronal lineage (Azizi et al., 1998). In utero transplanted MSC showed long term engraftment and multi-potential differentiation in a site-specific manner, acquiring the phenotype of the tissue in which they engraft (Liechty et al., 2000). Interestingly, some degree of plasticity has also been observed for mature cells within the mesenchymal lineage as chondrocytes as well as adipocytes transdifferntiated into osteoblasts (Kahn and Simmons, 1977; Bennett et al., 1991).

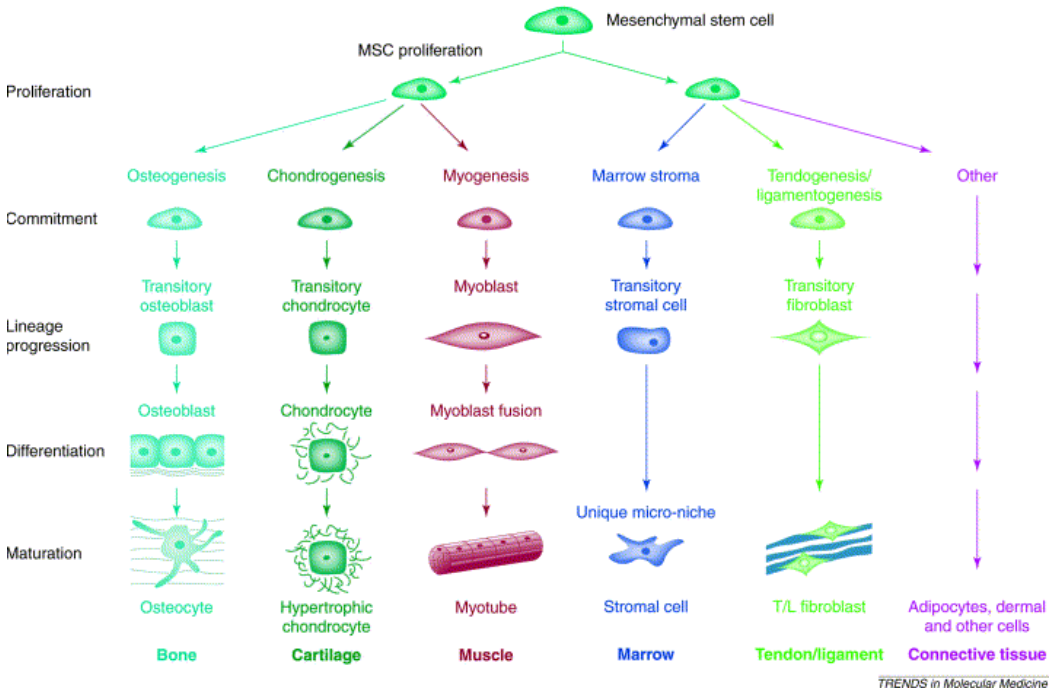

Figure 3 | Mesenchymal Stem Cell Differentiation. A simplified scheme of the stepwise cellular transitions from the mesenchymal stem cell (MSC) to the differentiated cells of the mesenchym tissues. The individual lineage pathways are arranged from left (best understood) to right (least understood); the osteogenic (Bruder et al., 1994) and chondrogenic (Boyan and Caplan, 1994) pathways are based on detailed experimental information (adapted from Caplan and Bruder, 2001).

Bone Morphogenetic Proteins

The Bone Morphogenetic[2] Proteins (BMPs) are a group of secreted bone-inducing proteins belonging to the transforming growth factor-β superfamily of growth factors. This family is composed of more than 40 related peptides. All BMPs (except BMP-1 which does not belong to the TGF-β superfamily) share significant sequence homology in the carboxy-terminal region with a conserved pattern of seven cysteine residues. Like other TGF-β superfamily members, BMPs are synthesized as precursor forms and are cleaved at the C-terminal region to yield mature proteins. Active BMP exist as dimers formed by disulfide bond bridging (Sampath et al., 1990).

Although it is not clear what are the specific roles of each of the BMPs knockout studies revealed that some of the BMPs are critical for normal development (including BMP-2 (Zhang and Bradley, 1996), BMP-4. (Winnier et al., 1995))

BMPs activate a signal transduction cascade by binding to receptors that activate the SMAD pathway leading to gene transcription (see Figure 4). BMP binds to a complex of two different types of serine/theronine kinase receptors: type I BMP receptrors, BMPR-1A and BMPR1-B, and type II receptors. After ligand binding, the type II receptor phosphorylates the type I receptor. The activated type I receptor phosphorylates a member of the SMAD family (SMAD 1,5, or 8) of intracellular proteins which then forms hetrodimeric complexes with SMAD 4. This complex directly regulates transcription. SMAD 6 and SMAD 7 are inhibitors of the TGF-β/BMP pathway.

[2] A morphogen is a secreted soluble molecule influencing differentiation of cells. A morphogen may specify more than one cell type by forming a concentration gradient.

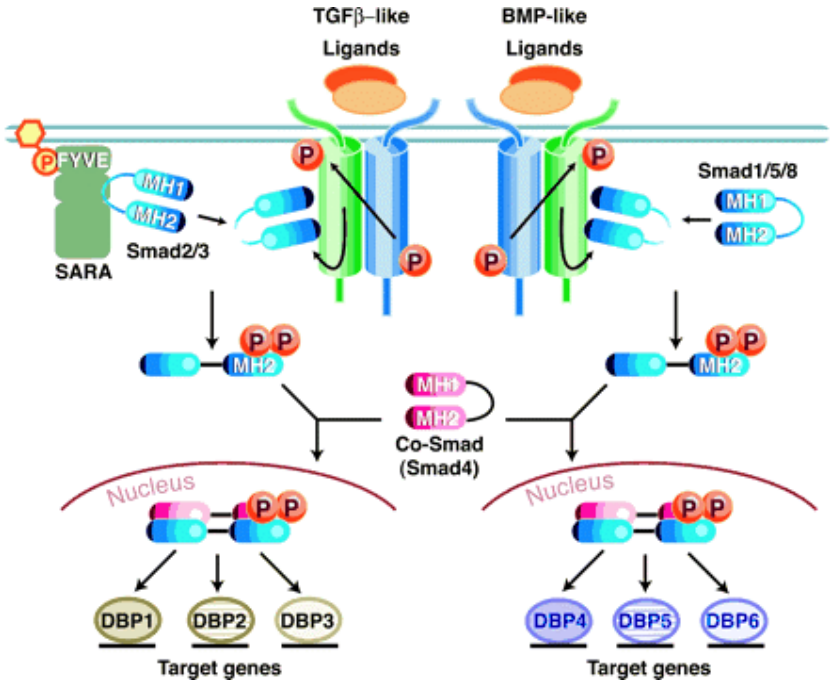

Figure 4 | TGF-β family ligands activate the Smad signal transduction cascade. The Smad pathway activated by TGF-β superfamily ligands. The ligand binding occurs with a complex of type I and type II receptors, type II receptor then activates type I receptor, by phosphorylating it on a particular serine or threonine residue. The phosphorylated type I receptor protein can now phosphorylate the Smad proteins. Those receptors that bind TGF-β family proteins or members of the activin family phosphorylate Smads 2 and 3. Those receptors that bind to BMP family proteins phosphorylate Smads 1, 5 and 8. These Smads can complex with Smad 4 to form active transcription factors that together with different DNA binding proteins transcribe target genes (adapted from Attisano and Wrana, 2002).

The BMP's were discovered (Wozney et al., 1988; Wang et al., 1988) after their isolation from demineralized bone matrix (DBM), which was shown to induce ectopic[3] bone formation in rodents (Urist, 1965; Van de Putte and Urist, 1965). The cellular events observed after DMB implantation mimic the endochondral bone formation process of embryonic development and fracture repair.

In vitro studies of BMP show that multipotent cells taken from animal or human marrow, do in fact respond to various BMP. Rat osteoprogenitor and ostoesarcoma (bone cancer) cells showed osteogenic differentiation, after exposure to recombinant human BMP-7 (Maliakal et al., 1994; Asahina et al., 1993). In other studies,

[3] Ectopic- A biological event or process (such as bone formation) that occurs in an abnormal location or position within the body.

cells from murine mesenchymal progenitor cell-line, C3H10T1/2, differentiated into osteoblasts after treatment with recombinant human BMP-2 (Yamaguchi et al., 1991; Katagiri et al., 1990). This osteoinductive behavior of the BMPs has been observed in numerous other studies as well (Wang et al., 1993; Chaudhari et al., 1997; Yamaguchi et al., 1996; Ahrens et al., 1993; Rickard et al., 1994; Balk et al., 1997; Rickard et al., 1994; Hanada et al., 1997). However, it seems that this activity is restricted to immature and multipotent cells, and does not affect mature cells (Katagiri et al., 1990; Yamaguchi et al., 1991; Kim et al., 1997; Knutsen et al., 1993).

Bone Therapy using BMP2

The unique properties of BMP make them possible candidates for use in bone therapy in cases in need for bone reconstruction such as fracture healing, osteoprosis and osteogenesis imperfecta. The main idea is to administer BMP in order to induce differentiation of MSC into osteogenic cells that will form bone.

Systematic administration of BMP-2 to osteoporotic mice increased their levels of MSCs in a dose dependent fashion. MSC were observed to proliferate and differentiate at an increased rate. Furthermore, stimulation of cartilage formation was observed, together with an increase in bone volume, and a decrease in MSC apoptosis (Turgeman et al., 2002b).

Local administration of BMP-2 in various studies using different animal models induced bone formation in-vivo (Wozney et al., 1988; Wozney et al., 1990; Volek-Smith and Urist, 1996). Furthermore, direct delivery of BMP-2 to areas of bone defects has been shown to induce *in vivo* bone healing (Marden et al., 1994; Kenley et al., 1994; Yasko et al., 1992; Lee et al., 1994; Kirker-Head et al., 1995; Toriumi et al., 1991; Sandhu et al., 1996; Zegzula et al., 1997). This therapeutic approach has already been tried out in human clinical trials for bone reconstruction, administering BMP-2 locally using

collagen sheets as carriers[4] (Boyne et al., 1997; Howell et al., 1997) . Unfortunately, the results of these studies were inconsistent, and varied depending on the location of the defect and on the individual patient (van den Bergh et al., 2000; Groeneveld and Burger, 2000).

A different therapeutic approach using BMPs for bone regeneration, which is currently being a subject of active research, is that of gene therapy (for a review see Alden et al., 2002). In general, the aim of gene therapy is to deliver a therapeutic gene, in our case BMP-2, into a target tissue. The advantages of such an approach compared to direct BMP application, is the long-term expression of BMP and the possibility to regulate its expression.

When considering gene therapy for bone repair, there are two main options: (i) direct gene delivery, *in vivo*, using viral or non-viral vectors, or (ii) *ex-vivo* cell mediated gene therapy. In *ex vivo* gene therapy, cells are removed from the patient for transfection with copies of the therapeutic gene. The treated cells are then returned to the patient.

Here we will focus on cell mediated gene therapy using MSCs genetically engineered to express the BMP-2 gene.

Using Genetically Engineered MSCs for Cell Mediated Gene Therapy

In *ex vivo* cell mediated gene therapy, cells are used for the delivery of genes into the tissues of interest. Usually in a typical gene therapy procedure, a functional copy of a gene is delivered into a tissue containing a defective gene. This is a different situation, in this case the cells have a functional intrinsic copy of the gene, but its regulation cannot be controlled externally. Therefore, in our case, the cells are engineered in such a way that the inserted copy of the gene of interest is attached to a regulatory element, that differs from that of the intrinsic gene, and therefore it's expression can be controlled from the outside (see The Experimental System, experiment outline).

[4] Carrier- a material that imparts the cells physical confinement or localization to the area of interest.

An advantage of using MSCs in bone gene therapy is that MSCs are able not only to deliver the gene, but also to participate in the process of the formation of the bone. Indeed, even unmodified MSCs can induce bone formation *in vivo* following transplantation in ectopic sites, and segmental bone defects (Bruder et al., 1998; Mankani et al., 2001) . However, this activity is dependent on the type of carrier[5] used, the carrier being osteo-inductive (inducing bone) and non-biodegradable. Genetically engineering of MSCs to express BMP-2 may perhaps elicit the osteogenic potential of the MSCs, regardless of the carrier type.

The use of MSCs in gene therapy has been a subject of active research (for reviews see Turgeman et al., 2002a; Pelled et al., 2002). Various studies have shown that MSCs were able to express, *in vivo*, genes that were transduced to them *ex-vivo* (Nolta et al., 1994; Bartholomew et al., 2001; Allay et al., 1997). Lou et al. showed that autologous MSCs that were transfected with recombinant human BMP2 (rhBMP2) *ex vivo*, were able to form bone in a spinal fusion model[6] *in vivo* (Lou et al., 1999). Bone formation was also induced using the same methods on the C3H10T1/2 mouse MSC cell line.

Further studies have shown that other osteoproginator cell lines transduced with rhBMP2 were able to induce bone formation in ectopic sites and to heal femoral defects in mice (Lieberman et al., 1998; Engstrand et al., 2000).

Gazit et al. (Gazit et al., 1999) raised the hypothesis, that genetically engineered mesenchymal stem cells, expressing the rhBMP-2 gene, display both paracrine and autocrine mechanisms[7] (Gazit et al., 1999 and Figure 5). This hypothesis was supported by three different studies carried by Gazit et al., using genetically engineered MSCs

[5] Carrier- a material that imparts the cells physical confinement or localization to the area of interest.

[6] Spinal fusion model- fusion of two parts of adjacent vertebrae induced by transplantation of cells that form bone in that site.

[7] When proteins synthesized by one cell can diffuse over small distances to induce changes in other cells, the event is called a **paracrine interaction**. **Autocrine** regulation occurs when the same cells that secrete paracrine factors also respond to them. In this case, the cell synthesizes a molecule for which it has its own receptor.

expressing rhBMP2 and Lac-Z marker[8] gene. In the first study a MSC clone constitutively expressing rhBMP2 and Lac-Z was used. The cells were transplanted in a murine radial segment defect[9], and a significant increase in bone formation was observed (Gazit et al., 1999). Notably, the engineered MSCs induced more bone healing compared to a non-MSC clone also expressing rhBMP-2, but possessing only a paracrine effect. In the second study, the MSCs were engineered to express rhBMP2 under external regulation (Moutsatsos et al., 2001). The external activation of rhBMP2 induced massive bone formation in radial segment defect, resulting in gap bridging. Furthermore, more bone was induced *in vivo* by the engineered MSCs than by rhBMP2 protein implants. In a more clinical oriented model, human MSC were engineered to express rhBMP2 and LacZ. In this study, the cells were transplanted in radial segment defects in mice and showed a complete gap bridging forty-five days after transplantation (Turgeman et al., 2001). All these studies demonstrated that the genetically engineered MSCs are able to engraft, and to differentiate into cartilage and bone cells. Furthermore, the MSCs were spotted, using Lac-Z, in every stage of the bone development but unmarked cells from the host were also observed to participate in the repair process. Thus, one can infer that rhBMP2 expression led to the recruitment and differentiation of the host cells through a paracrine effect, and that the differentiation of the genetically engineered MSCs is through an autocrine mechanism.

Taken together, all these studies have shown that genetically engineered MSCs have the capacity to engraft, differentiate and also to recruit host cells in order to grow bone. As such, genetically engineered MSCs are a promising therapeutic tool not only for fracture healing, but also for many other clinical situations, such as

[8] Lac-Z- a gene used as marker, colors the cells in blue.

[9] Radial segment defect- procedure applied to the Radius bone (one of the long bones of the arm) in which a segment (a piece) of the bone is cut. The size of the segment is too large for the defect to heal by itself, without any external intervention (resulting in a fracture referred to as non-union fracture).

osteoprosis, Osteogenic Imperfecta, and more, which require bone regeneration.

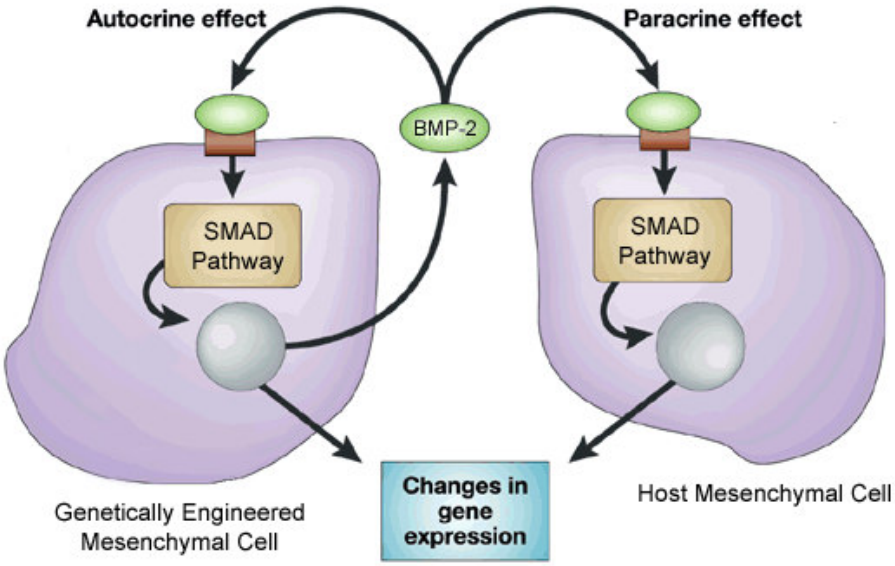

Figure 5 | Model proposed by Gazit et al. the effects of therapy based on genetically engineered MSCs. The diagram represents the paracrine and autocrine mechanisms for the control of bone regeneration. Following transplantations of Genetically Engineered Mesenchymal Stem Cells, rhBMP-2 expression enhances host mesenchymal cell differentiation (paracrine mechanism) and also the differentiation of the transplanted cells themselves (autocrine mechanism).

Our goal is to investigate the process by which genetically engineered rhBMP2 expressing adult MSCs form bone *in vivo.* Previous data has shown that these cells form bone *in vivo* through the endochondral pathway (Turgeman et al., 2001). We would like to characterize this process using DNA chips (GeneChip™ technology) and to pinpoint genes involved in the osteogenic process that may later serve as candidates for skeletal therapy.

The Experimental System

All the experiments were designed and preformed by Hadi Aslan at Prof. Gazit's lab at The Hebrew University-Hadassah Medical center Ein Kerem in Jerusalem.

General Outline of the System

Murine cell line MSCs were genetically engineered to carry the rhBMP2 and Lac-Z genes, the clone that was generated was named C9. The expression of the rhBMP2 is controlled by a tet-off system.

The tet-off system puts the gene of interest under the control of a transactivator (tTA) which is constitutively expressed. Thus the gene of interest is also constitutively expressed, unless the transactivator is deactivated by dox (Doxycycline, a derivative of tetracycline). Thus, the expression of rhBMP2 is controlled exogenously: when exposed to doxycycline the cells do not express rhBMP2, and when the cells are not exposed to doxycycline rhBMP2 is expressed.

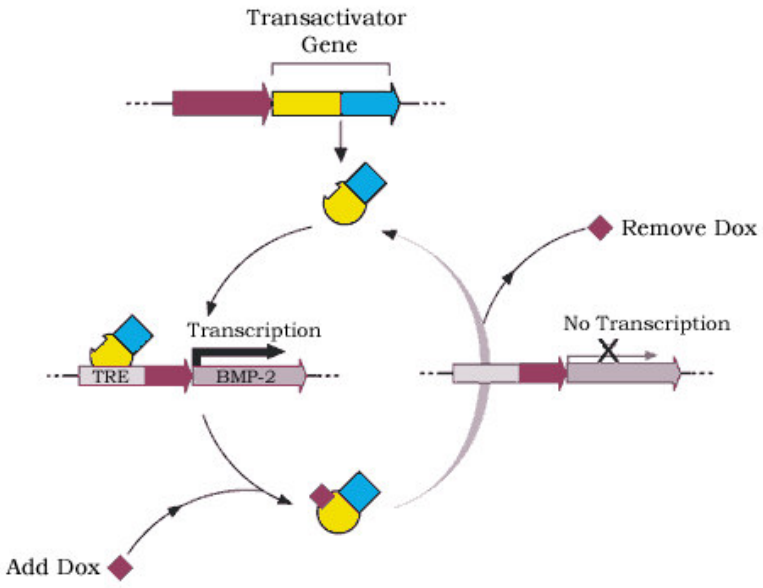

Figure 6 | Tet-Off system. A transactivator gene with a constitutive promoter is inserted into the cells along with the BMP-2 gene, having a Tet Response Element (TRE) in its promoter. The transactivator transcribes from the TRE only in the absence of Dox. Addition of Dox eliminates gene transcription.

The cells were transplanted subcutaneously into mice. Mice were divided into two groups: (i) Baseline (BAS): a control group which received DOX in the drinking water, and therefore the transplanted cells did not express rhBMP2, and (ii) Experiment (EXP): a group which was not exposed to DOX and therefore the transplanted cells did express rhBMP2.

Bone Formation

Harvests from both groups were taken at seven different time points: days, 1, 3, 5, 7, 10, 15, 20-post transplantation. Five different harvests were taken for each time point in each group.

The C9 cells implanted in the EXP group were found to undergo a process resembling endochondral bone formation (Figure 7). Histological analysis of the forming tissue, showed a dynamic diffrentiation process which can be roughly divided into three major phases: (i) the "mesenchymal phase" days 1-5, in which mainly mesenchymal cells are apparent, (ii) the "cartilage phase" days 7-10, cartilage then appears and matures towards its transition to bone on day 10, and (iii) the "mature phase" where mature cartilage and bone can be seen (figure 7).

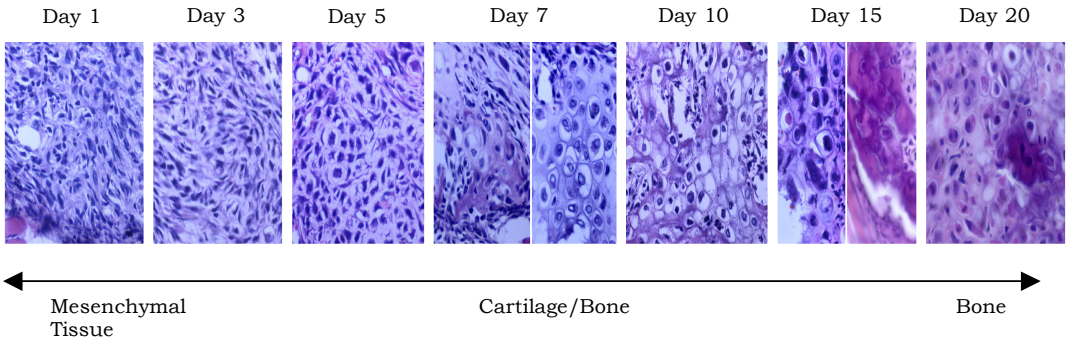

Figure 7 | Osteogenic differentiation of rhBMP2 engineered MSCs in vivo through the endochondral pathway.

DNA Chips

RNA extractions were labeled and hybridized to two different Affymetrix chips : (i) Mu1a: measuring about 1200 genes, there were 70 chips (7 time points, each time point 10 chips: 5 BAS and 5 EXP), and (ii) U74: measuring about 38,000 genes, here also, there were 70 chips.

The DNA chip raw data was given to us for further analysis.

Results

The Data Set

Samples were taken from mice at seven different timepoints. For each time point 5 samples were taken from the control group, and 5 from the EXP group. The experiments were done using two different Affymetrix chips: the U74 chip, containing about 38,000 genes and the Mu1a chip containing 1200 genes. The U74 was used in two versions: the first set of 7 time points, that is, 1 EXP and 1 BAS chip for each time point, was done using the old version of the chip (14 chips). The rest of the measurements, i.e. four sets of seven time points each, were done using the improved version of the U74 chip (56 chips). Unfortunately, the first version of the U74 chip contained about 12,000 defective probes, leaving us with about 26,000 genes. Thus, we had three different data sets: For the U74 chip we had a matrix of n_S=*70* columns (experiments) and 26,000 rows (genes on the chip) (we will refer to it as U74 v1 dataset), or a matrix of n_S=*56* columns (experiments) and 38,000 rows (genes on the chip) (we will refer to it as U74 v2 datset). For the Mu1a chip we had a matrix of n_S=*70* columns (experiments) and 1200 rows (genes on the chip). The Mu1a matrix was later reduced in size since one set of 7 time points (14 chips) looked extremely different than the others and was removed, leaving us with a matrix of n_S=*56.*

Preprocessing and filtering

Affymetrix controls were removed. We filtered the data as described in the methods section, using $\sigma_T = 0.6$ for the Mu1a dataset. This filtering left us with about 400 genes. The U74 v2 dataset was filtered using $\sigma_T = 0.8$ and $\sigma_T = 0.85$, leaving us with about 3700 genes, and 3000 genes respectively. The U74 v1 dataset was filtered using $\sigma_T = 0.8$, leaving us with about 3000 genes. There was no need to

threshold the data since the expression values were given as frequencies (see Methods chapter 5).

Combined Supervised/Unsupervised Analysis

At first, we used the Two Way Anova test (see Methods chapter 5) in order to find genes differentially expressed between the Baseline and Experiment samples over time. The two dimensions that were used for the test are: (i) Type of sample, that is, Bas or Exp, and (ii) the time point (day 1, day 3 etc.). The type of genes that one would expect to retrieve using such a test are genes whose behavior over time is different between Baseline samples and Experiment samples.

The FDR threshold was set to a very stringent value of 1%, meaning that the expected value of the genes falsely discovered is only 1%.

The second step was unsupervised; we clustered the genes that were identified as differentiating, in order to group them into clusters with similar behaviors. We used the SPC clustering algorithm.

Using this method we received some very interesting clusters of genes, which were reproduced (with some variation), in the analysis of all datasets. We chose to focus on four of these clusters.

Cluster 1: immunoglobulin chains

This cluster, having a very distinctive profile, consistently appeared. Most of the 44 genes that constitute the cluster are of unknown function, but almost all the genes that are known are genes of immunoglobulin chains (Table 1, see table 1 in Appendix C for a full list of genes). These genes are highly expressed in days 15 and 20 of the experiment and are expressed at low levels both in the controls and in the earlier experiment time points (Figure 8).

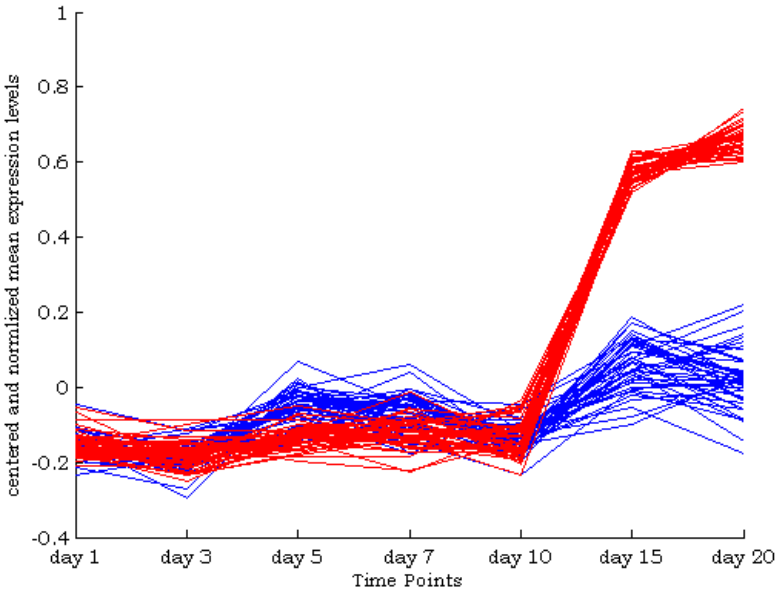

Figure 8 | Profiles of the "Immunoglobulin Cluster". The mean expression level of the genes in each group of 5 samples of the same time point is plotted. Blue lines are BAS mean expression values and red are EXP mean expression values. .

Table 1 | Identified genes of Cluster 1

Probe Set ID	Accesion	Gene Symbol	Description
92470_f_at	AF065324	AI893585	expressed sequence AI893585
102823_at	X67210	AU044919	expressed sequence AU044919
101870_at	V00793	Igh-4	immunoglobulin heavy chain 4 (serum IgG1)
101871_f_at	X88902	Igh-4	immunoglobulin heavy chain 4 (serum IgG1)
101745_f_at	X94422	Igh-6	immunoglobulin heavy chain 6 (heavy chain of IgM)
100299_f_at	U55576	Igk-V28	immunoglobulin kappa chain variable 28 (V28)
101331_f_at	U55641	Igk-V28	immunoglobulin kappa chain variable 28 (V28)
93213_at	AB007986	Igk-V28	immunoglobulin kappa chain variable 28 (V28)
_101640f_at	M15593	Igk-V28	
_101718f_at	U19315	Igk-V28	
_93213at	AB007986	Igk-V28	
_93227f_at	Z70661	Igk-V28	
100682_f_at	U62386		Mus musculus immunoglobulin heavy and light chain variable region
102843_s_at	J00475		Mus musculus, Similar to immunoglobulin heavy chain 1 (serum IgG2a)
93927_f_at	L33954		Mus musculus, Similar to immunoglobulin heavy chain 1 (serum IgG2a)
100721_f_at	U10410	LOC56304	recombinant antineuraminidase single chain Ig VH and VL domains
111469_at	AA794875	2010001M09Rik	RIKEN cDNA 2010001M09 gene

Cluster 2: Degradation Enzymes

This cluster behaves in a similar way to the previous one, although the increase in expression starts earlier, already obvious at day 10, and there is a slight decrease of expression on day 20 (Figure 9). This cluster contains six genes; four of them are degradation enzymes (LIPC, MMP9, CTSK, FAP), one encodes a protein that constitutes the bone matrix (BGLAP-RS1) and the last gene (CKB) encodes a protein involved in energy transfer (Table 2).

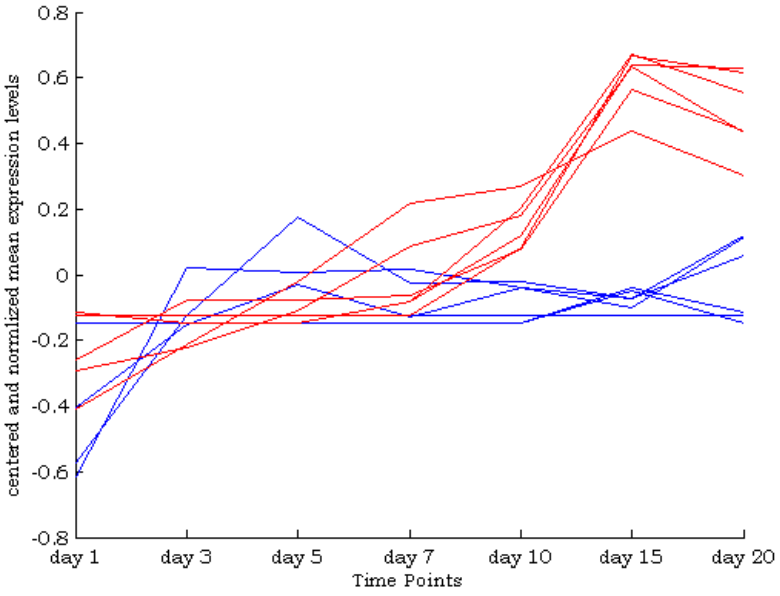

Figure 9 | Profiles of the "Degradation Cluster". The mean expression level of the genes in each group of 5 samples of the same time point is plotted. Blue lines are BAS mean expression values and red are EXP mean expression values.

Table 2 | Genes of Cluster 2

Gene ID	Symbol	Description
X58426_at	Lipc	lipase, hepatic
L24430_x_at	Bglap-rs1	bone gamma-carboxyglutamate protein, related sequence 1
Z27231_at	Mmp9	matrix metalloproteinase 9
M74149_x_at	Ckb	creatine kinase, brain
X94444_at	Ctsk	cathepsin K
Y10007_at	Fap	fibroblast activation protein

Cluster 3: Collagens, Proteolytic Enzymes, and WNT related proteins

This cluster contains about 80 genes, some of which belong to very distinct functional groups. These groups consist of a large number of

collagen genes and extra-cellular matrix proteins, genes related to the WNT pathway, various bone related genes (such as Bambi, a bmp and activin inhibitor, RUNX1- a transcription factor maybe involved in bone development, and proteolytic degradation enzymes. In addition to these, there are a few proteoglycans which are a known part of the ECM, but also play a role in signal transduction of various pathways including the WNT pathway. The genes of the cluster increase in expression during early EXP, peaking around day 10 and then decreasing in expression during the late stage of the experiment, while maintaining a relatively low and constant expression level in BAS (Figure 10, table 3, see Appendix C table 2 for full list of genes)

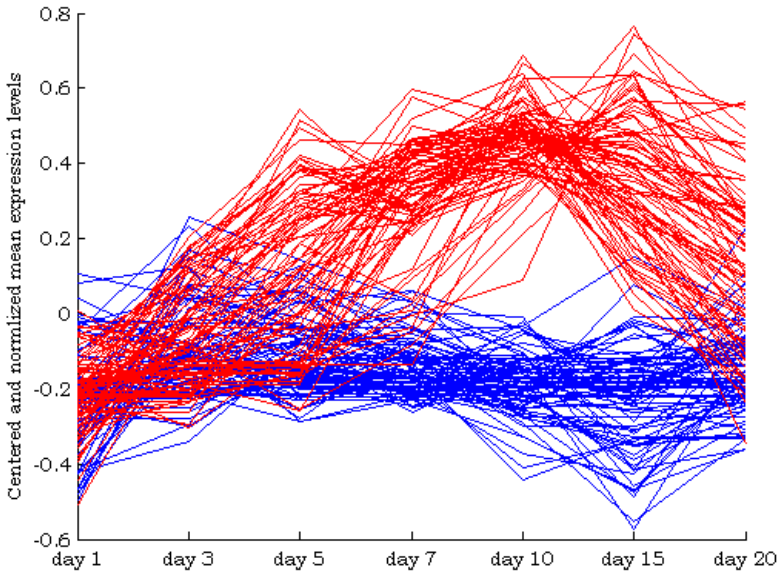

Figure 10 | Profiles of the "Collagens, Proteolytic and WNT Cluster". The mean expression level of the genes in each group of 5 samples of the same time point is plotted. Blue lines are BAS mean expression values and red are EXP mean expression values.

Table 3 | Categorized Genes of Cluster 3

Cell Adhesion,Collagen, and Extracellular Matrix

Probe ID	Accession	Symbol	Description
100308_at	X66976	Col8a1	procollagen, type VIII, alpha 1
102070_at	AW212495	Col9a3	procollagen, type IX, alpha 3
103616_at	AF100956	Col11a2	procollagen, type XI, alpha 2
104483_at	L12215	Col9a1	procollagen, type IX, alpha 1
100476_at	L20232	Ibsp	integrin binding sialoprotein
100481_at	D38162	Col11a1	procollagen, type XI, alpha 1

102261_f_at	U30292	Col13a1	procollagen, type XIII, alpha 1
163036_at	AV243867	Col9a1	procollagen, type IX, alpha 1
163627_f_at	AI847474	Cthrc1	collagen triple helix repeat containing 1
92207_at	U08210	Eln	Elastin
100336_s_at	L24431	Bglap2	bone gamma-carboxyglutamate protein 2
99903_at	AJ242625	Dmp1	dentin matrix protein 1
93369_at	AB007848	Omd	Osteomodulin
116382_at	AW046112	Nrp2	neuropilin 2

Proteoglycans

Gene ID	Accession	Symbol	Description
93136_at	D78274	Dspg3	dermatan sulphate proteoglycan 3
104205_at	L07049	Agc1	aggrecan 1
160561_at	M34094	Mdk	Midkine
108045_at	AA798520	Lepre1	leprecan 1

Wnt pathway related

Gene ID	Accession	Symbol	Description
103513_at	M89799	Wnt5b	wingless-related MMTV integration site 5B
93188_at	AJ243964	Dkk3	dickkopf homolog 3 (Xenopus laevis)
106051_at	AI838192	Dkk3	dickkopf homolog 3 (Xenopus laevis)
104672_at	U68058	Frzb	frizzled-related protein
163423_at	AW107799	Wif1	Wnt inhibitory factor 1
166263_at	AV346158	Frzb	frizzled-related protein
116807_at	AI846342	Mfrp	membrane-type frizzled-related protein

Proteases

Gene ID	Accession	Symbol	Description
97941_at	Y12582	Capn6	calpain 6
97942_g_at	Y12582	Capn6	calpain 6
99958_at	J05177	Mcpt2	mast cell protease 2
97109_at	L00653	Mcpt7	mast cell protease 7
100484_at	X66473	Mmp13	matrix metalloproteinase 13
99957_at	X72795	Mmp9	matrix metalloproteinase 9
165565_r_at	AI481262	Pcolce2	procollagen C-endopeptidase enhancer 2
108365_at	AA981778	Pcolce2	procollagen C-endopeptidase enhancer 2

Ligands (Growth factors), cell growth

Gene ID	Accession	Symbol	Description
103244_at	AB009245	Scgf	stem cell growth factor
165981_i_at	AV140849	Lect1	leukocyte cell derived chemotaxin 1
114346_at	AA960023	Il17b	interleukin 17B

Bone development

Gene ID	Accession	Symbol	Description
98482_at	X78936	Pthr	parathyroid hormone receptor
92796_at	J02980	Akp2	alkaline phosphatase 2, liver
162531_at	AW048375	Bambi	BMP and activin membrane-bound inhibitor, homolog (Xenopus laevis)

Transcription Factors			
Gene ID	Accession	Symbol	Description
133900_at	AI481773	Runx1	runt related transcription factor 1
135401_at	AW214326		ESTs, Highly similar to SOX8_MOUSE Transcription factor SOX-8 [M.musulus]

Cluster 4: Osteogenesis, Adhesion , Angiogenesis and Frizzeled

This cluster contains, in addition to collagens and proteases, angiogenesis inducing molecules such as TGFβ1, TGF β3 and CYR61 and also receptors for angiogenesis inducing growth factors, such as PDGF (patlet derived growth factor) receptor. In addition to that, this cluster includes many adhesion related genes such as cadeherins and integrin ligands which mediate cell-cell adhesions and adhesion to matrix proteins respectively, and a Gap junction protein, GJA1, involved in cell-cell communication. Genes involved in osteogenesis including BMP2/4 receptor, BMP1, and SMAD6 (MADH6) and follistatin which are inhibitors of BMP signaling. This set of genes also contains two different frizzeled receptors for Wnt. RUNX2, a transcription factor involved in bone differentiation is also present. In this cluster the controls show a decrease in expression with time, whereas the Exp exhibit a mild peak at day 10 (Figure 11, table 4, see Appendix C table 3 for full list of genes).

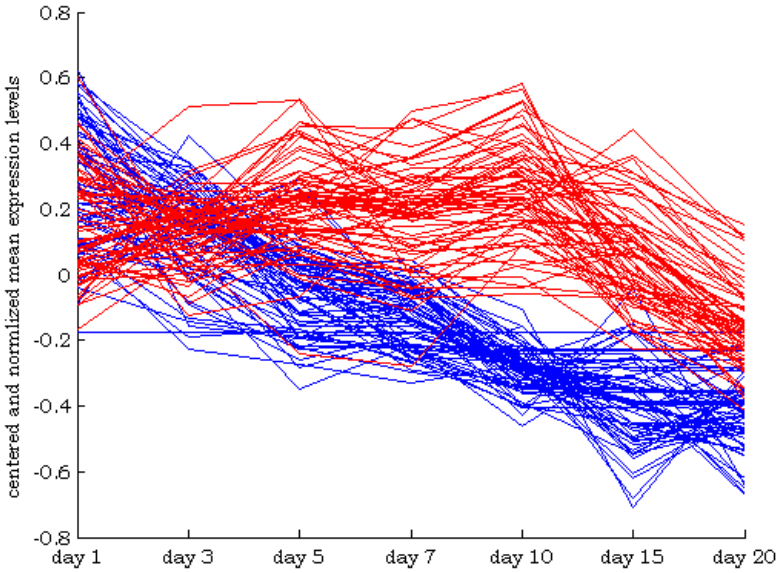

Figure 11 | Profiles of the "Osteogenesis, Adhesion ,Angiogenesis and Frizzeled Cluster". The mean expression level of the genes in each group of 5 samples of the same time point is plotted. Blue lines are BAS mean expression values and red are EXP mean expression values.The profiles of the genes belonging to the cluster. Blue lines are BAS mean expression values and red are EXP mean expression values.

Table 4 Categorized Genes of cluster 4

Angiogenesis and ligands (growth factors) and their receptors

Gene ID	Symbol	Description
M32490_at	Cyr61	cysteine rich protein 61
M32745_at	Tgfb3	transforming growth factor, beta 3
X04367_at	Pdgfrb	platelet derived growth factor receptor, beta polypeptide
AF056187_at	Igf1r	insulin-like growth factor I receptor
M70642_at	Ctgf	connective tissue growth factor
M13177_at	Tgfb1	transforming growth factor, beta 1
L08394_at	Btc	Betacellulin, epidermal growth factor family member

Cell cycle and apoptosis

Gene ID	Symbol	Description
Z26580_at	Ccna2	cyclin A2
Y13087_at	Casp6	caspase 6
U24173_at	Cdkn1a	cyclin-dependent kinase inhibitor 1A (P21)

Wnt Pathway Related

Gene ID	Symbol	Description
AF139183_at	Fzd2	frizzled homolog 2 (Drosophila)
U43321_at	FZD8	frizzled homolog 8 (Drosophila)
AF126063_at	Wisp2	WNT1 inducible signaling pathway protein 2

Collagens, Extracellular Matrix

Gene ID	Symbol	Description
M18933_at	Col3a1	Procollagen, type III, alpha 1
AB009993_at	Col5a1	Procollagen, type V, alpha 1
M65161_x_at	Col2a1	Procollagen, type II, alpha 1
U25652_at	Col12a1	Procollagen, type XII, alpha 1
L02918_at	Col5a2	Procollagen, type V, alpha 2
D17511_at	Col9a1	Procollagen, type IX, alpha 1
AF080572_x_at	Plod2	Procollagen lysine, 2-oxoglutarate 5-dioxygenase 2
M65142_x_at	Lox	lysyl oxidase

Proteoglycans

Gene ID	Symbol	Description
X15487_x_at	Sdc1	syndecan 1
D89571_at	Sdc4	syndecan 4
X70296_at	Serpine2	serine (or cysteine) proteinase inhibitor, clade E, member 2
X53928_at	Bgn	Biglycan

Proteases

Gene ID	Symbol	Description
U54984_at	Mmp14	matrix metalloproteinase 14 (membrane-inserted)
M84324_at	Mmp2	matrix metalloproteinase 2
Z12604_at	Mmp11	matrix metalloproteinase 11
AB008548_at	Pcolce	Procollagen C-proteinase enhancer protein

Adhesion

Gene ID	Symbol	Description
AF112151_at	Fbln5	fibulin 5
M27129_x_at	Cd44	CD44 antigen
X07233_at	Ncam1	neural cell adhesion molecule 1
AB008811_x_at	Cdh2	cadherin 2
X61576_at	Gja1	gap junction membrane channel protein alpha 1
D21253_at	Cdh11	cadherin 11
X56304_at	Tnc	tenascin C

BMP-Related

Gene ID	Symbol	Description
AF010133_at	Madh6	MAD homolog 6 (Drosophila)
Z29532_at	Fst	Follistatin
L25602_at	Bmp2	bone morphogenetic protein 2
L24755_at	Bmp1	bone morphogenetic protein 1
U04673_at	Bmpr1a	bone morphogenetic protein receptor, type 1A

Transcription Factors

Gene ID	Symbol	Description
M98339_at	Gata4	GATA binding protein 4
D14636_at	Runx2	runt related transcription factor 2

Activins, follistatin

Gene ID	Symbol	Description
X69619_x_at	Inhba	inhibin beta-A
Z29532_at	Fst	Follistatin
X69620_at	Inhbb	inhibin beta-B

Inflammation

Gene ID	Symbol	Description
M72394_at	Pla2g4a	Phospholipase A2, group IVA (cytosolic, calcium-dependent)
M64291_at	Ptgs2	prostaglandin-endoperoxide synthase 2

Unsupervised Analysis

CTWC was applied to the data, and the resulting stable clusters were further searched in order to find clusters of genes that separate the BAS and EXP samples. One gene cluster (appearing consistently in all U74 datasets analyzed) was found to separate the samples into BAS and EXP, and also order the EXP samples chronologically (Figure 12) (denoted cluster 5).

The genes of this cluster have low expression in BAS and early EXP, but their expression values rise at middle and late EXP. Furthermore, one can see two kinds of genes: genes that go up early (days 3, 5), and go down at the late stages of the experiment (days 15, 20)(genes 1-33 in figure 12), and genes that go up at the mid-stage of the experiment (days 7, 10), and are still elevated in the late stages of the experiment (genes 34-60 in figure 12).

Not surprisingly, this cluster contains many genes that were found in the supervised analysis (Table 5, see Appendix C table 4 for a complete list of genes in the original order). One of the genes that was found in this cluster and was not found in the supervised analysis is Tbx2, a transcription factor that is known to be involved in mesoderm differentiation and limb patterning.

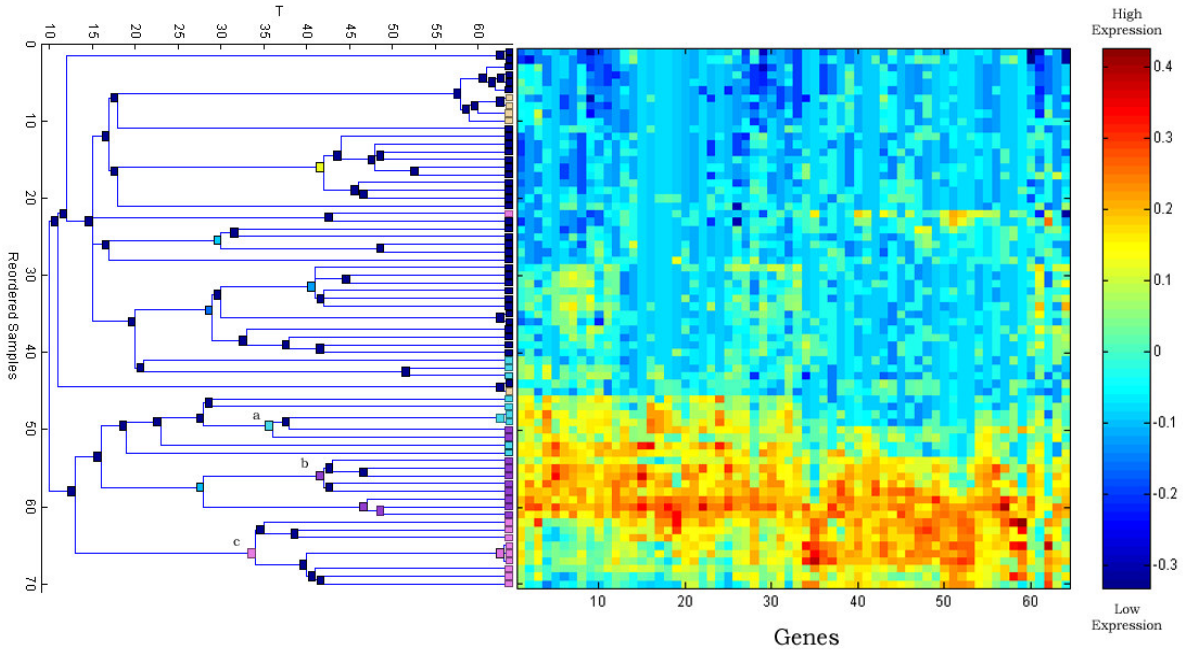

Figure 12 | Clustering of the samples according to the genes of cluster 5. Bas samples are dark blue; Exp samples: day 1: white, days 3 and 5 (early experiment) are turquoise, days 7 and 10 (mid experiment) are purple and days 15 and 20 (late experiment) are pink. The Exp samples are ordered chronologically, from early experiment to late. Cluster 'a' is mostly early Exp, cluster 'b' is mostly mid Exp, and cluster 'c' is late Exp.

Table 5 | Categorized Genes of Cluster 5

Collagens, Extracellular Matrix

Probe ID	Accession	Symbol	Description
93369_at	AB007848	Omd	Osteomodulin
100308_at	X66976	Col8a1	procollagen, type VIII, alpha 1
103616_at	AF100956	Col11a2	procollagen, type XI, alpha 2
100481_at	D38162	Col11a1	procollagen, type XI, alpha 1
92506_at	AF098460	Crtl1*	cartilage link protein 1
102261_f_at	U30292	Col13a1	procollagen, type XIII, alpha 1
99903_at	AJ242625	Dmp1	dentin matrix protein 1

Proteoglycans

Probe ID	Accession	Symbol	Description
104205_at	L07049	Agc1	aggrecan 1
93136_at	D78274	Dspg3	dermatan sulphate proteoglycan 3
108045_at	AA798520	Lepre1	leprecan 1

Proteases

Probe ID	Accession	Symbol	Description
108365_at	AA981778	Pcolce2	procollagen C-endopeptidase enhancer 2
97941_at	Y12582	Capn6	calpain 6
97942_g_at	Y12582	Capn6	calpain 6

99957_at	X72795	Mmp9	matrix metalloproteinase 9
97109_at	L00653	Mcpt7	mast cell protease 7
100484_at	X66473	Mmp13	matrix metalloproteinase 13

Adhesion

Probe ID	Accession	Symbol	Description
116382_at	AW046112	Nrp2	neuropilin 2
100336_s_at	L24431	Bglap2	bone gamma-carboxyglutamate protein 2
100476_at	L20232	Ibsp	integrin binding sialoprotein
94930_at	L07803	Thbs2*	Thrombospondin 2

Wnt Pathway Related

Probe ID	Accession	Symbol	Description
116807_at	AI846342	Mfrp	membrane-type frizzled-related protein
104672_at	U68058	Frzb	frizzled-related protein
106051_at	AI838192	Dkk3	dickkopf homolog 3 (Xenopus laevis)

Ligands (Growth Factors)

Probe ID	Accession	Symbol	Description
103244_at	AB009245	Scgf	stem cell growth factor
101918_at	AJ009862	Tgfb1*†	transforming growth factor, beta 1
114346_at	AA960023	Il17b	interleukin 17B

Bone Development

Probe ID	Accession	Symbol	Description
98482_at	X78936	Pthr	parathyroid hormone receptor
92796_at	J02980	Akp2	alkaline phosphatase 2, liver

Transcription Factors

Probe ID	Accession	Symbol	Description
133900_at	AI481773	Runx1	runt related transcription factor 1
92705_at	U15566	Tbx2*	T-box 2
135401_at	AW214326		ESTs, Highly similar to SOX8_MOUSE Transcription factor SOX-8 [M.musulus]

* Genes that were NOT found in cluster 3 of the supervised analysis(all others are found).

† This genes appeared in cluster 4 of the supervised analysis.

Discussion

Two Way ANOVA Combined with SPC provide a Powerful Analysis Tool.
We used two-way ANOVA in order to screen the genes that have a difference of behavior between BAS and EXP over the time. These genes were then clustered using SPC in order to assign them to groups with similar profiles.
Given the amount of meaningful results we received and their consistency throughout the datasets, we conclude that these methods combined provide powerful profile analysis tools for timecourse chip experiments.

The Differentiating Cells Secrete Structural Components, Growth Factors and Enzymes
The results show that many genes involved in bone matrix formation are upregulated. These include various collagens, proteoglycans which may also play a role in signaling (Perrimon and Bernfield, 2000), and adhesion molecules. In addition, various growth factors are transcribed, having roles in angiogenesis (vascular penetration), cell proliferation and differentiation. Finally, different degredation enzymes are produced in order to degrade unwanted cartilage and bone matrix and facilitate the release of various factors (such as SCF and TGF-β).

Activin and its Antagonists Are Expressed During Endochondral Bone Formation.
Activin is a TGF-β superfamily member, existing as covalently linked hetrodimers or homodimers of two closely related βA and βB chains. Activin is mostly known as a regulator in the reproductive endocrine system. However, activin seems to play a role also in the regulation of a variety of processes related to development differentiation, and growth, including skeletal development. Activin was also found to induce osteoblast formation in vitro (Gaddy-Kurten et al., 2002). Follistatin and BAMBI are activin binding proteins, both inhibiting

activin's interaction with its receptor. BAMBI and follistatin also interact with and inhibit the BMP2 signaling pathway. It has been shown that activin subunit βA and follistatin expression was upregulated in a mouse osteobalstic cell line in response to BMP2 exposure (Kearns and Demay, 2000). The authors suggested that activin expression is induced by BMP2, and its activity, and maybe BMP2 activity as well, is resricted by follistatin. Our results support this hypothesis and also add some new information, as we found upregulation of not only activin subunit βA but also subunit βB. In addition to that BAMBI, which is also an activin and BMP2 inhibitor, was upregulated. Taken together, it seems that BMP2 expression induces production of activin and its inhibitors, suggesting that activin and BMP2 function is modulated by these inhibitors during endochondral bone formation.

Immunogolublin mRNA at day 15 and 20 of the Experiment

One of our very stable clusters following the ANOVA analysis was a 40- gene cluster, genes having an almost identical behavior. These genes' expression levels rise at the late stages of the experiment, days 15 and 20. Only 16 of the cluster's genes are known genes; interestingly, 13 of these genes are related to immunoglobulin. These genes include heavy and light chains, and variable regions. A possible explanation for this observation is the formation of bone marrow (Wozney et al., 1990) and the subsequent arrival of B cells.

WNT signaling is involved in Endochondral bone Formation

A very interesting result is the upregulation of WNT pathway related proteins, including WNT5b, Frizzeled receptors and various inhibitors of the Wnt pathway, Dkk3, Wif and FRZB. Wnts are a group of secreted signaling glycoproteins that play an important role in many different fundamental processes, including embryonic axis specification, cell fate determination, differentiation and organogenesis (Cadigan and Nusse, 1997; Moon et al., 1997).

Wnt signaling is mediated through different intracellular pathways; the most characterized is the beta-catenin LEF/TCF. Wnt binding to the frizzeled receptor and also possibly a Lipoprotein-Receptor-related protein 5 (LRP5) coreceptor (Gong et al., 2001), results in the stabilization of beta-catenin, which enters the nucleus and interacts with transcription factors of the lymphoid enhancer binding factor (LEF) and T-cell transcription factor families resulting in gene expression (Figure 13). Secreted frizzeled proteins and other proteins, such as WIF and DKK act as inhibitors of the WNT pathway.

Various Wnts were found to be expressed in unique spatial and temporal patterns during skelatogenesis (Dealy et al., 1993; Parr and McMahon, 1995; Kengaku et al., 1998). In addition to that, two recent studies have implicated Wnt signaling in bmp-2 stimulating chondrogenesis (Fischer et al., 2002b; Fischer et al., 2002a) and ectopic endochondral ossification (Kitagaki et al., 2003). Frzb, a natural antagonist of certain Wnts, was found to be expressed during cartilage development and to regulate chondrocyte maturation and long bone development during limb skeletogenesis. Furthermore, a mutation in LRP5 (Gong et al., 2001) is a cause of a disease termed osteoprosis pseudoglioma syndrome, in which patients have very low bone mass. Thus, it seems that the WNT pathway does indeed play a role in bone development.

Our recent results further support these findings, and suggest the possibility of crosstalk between the BMP2 and WNT pathways. Our findings show that in addition to the expression of wnt5B and of the membrane FRZ receptors, three natural inhibitors of Wnts are expressed, suggesting complex interplay during endochondral ossification between Wnt and its inhibitors.

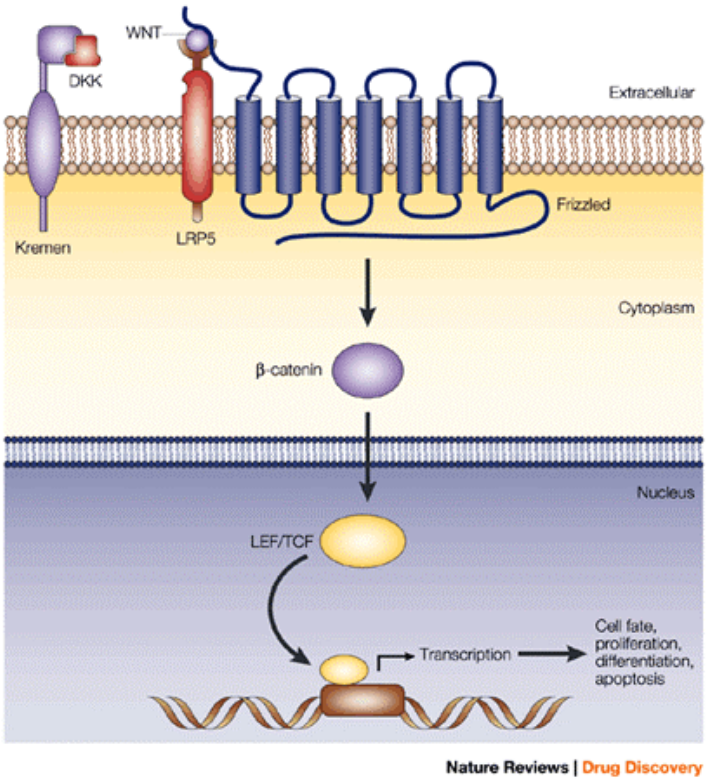

Figure 13 |WNT signaling pathway. Wnt interacts with the Frizzeled receptor and LRP5, resulting in cellular events leading to beta-catenin stabilization. Beta-catenin enters the nucleus, interacts with LEF/TCF transcription factors and activates specific target genes. DKK and Kerman functionally cooperate to inhibit WNT signaling through the LRP5 receptor.

Reference List

Ahrens,M., Ankenbauer,T., Schroder,D., Hollnagel,A., Mayer,H., and Gross,G. (1993). Expression of human bone morphogenetic proteins-2 or -4 in murine mesenchymal progenitor C3H10T1/2 cells induces differentiation into distinct mesenchymal cell lineages. DNA Cell Biol. *12*, 871-880.

Alden,T.D., Varady,P., Kallmes,D.F., Jane,J.A., Jr., and Helm,G.A. (2002). Bone morphogenetic protein gene therapy. Spine *27*, S87-S93.

Allay,J.A., Dennis,J.E., Haynesworth,S.E., Majumdar,M.K., Clapp,D.W., Shultz,L.D., Caplan,A.I., and Gerson,S.L. (1997). LacZ and interleukin-3 expression in vivo after retroviral transduction of marrow-derived human osteogenic mesenchymal progenitors. Hum. Gene Ther. *8*, 1417-1427.

Asahina,I., Sampath,T.K., Nishimura,I., and Hauschka,P.V. (1993). Human osteogenic protein-1 induces both chondroblastic and osteoblastic differentiation of osteoprogenitor cells derived from newborn rat calvaria. J. Cell Biol. *123*, 921-933.

Attisano,L. and Wrana,J.L. (2002). Signal transduction by the TGF-beta superfamily. Science *296*, 1646-1647.

Azizi,S.A., Stokes,D., Augelli,B.J., DiGirolamo,C., and Prockop,D.J. (1998). Engraftment and migration of human bone marrow stromal cells implanted in the brains of albino rats--similarities to astrocyte grafts. Proc. Natl. Acad. Sci. U. S. A *95*, 3908-3913.

Balk,M.L., Bray,J., Day,C., Epperly,M., Greenberger,J., Evans,C.H., and Niyibizi,C. (1997). Effect of rhBMP-2 on the osteogenic potential of bone marrow stromal cells from an osteogenesis imperfecta mouse (oim). Bone *21*, 7-15.

Bartholomew,A., Patil,S., Mackay,A., Nelson,M., Buyaner,D., Hardy,W., Mosca,J., Sturgeon,C., Siatskas,M., Mahmud,N., Ferrer,K., Deans,R., Moseley,A., Hoffman,R., and Devine,S.M. (2001). Baboon mesenchymal stem cells can be genetically modified to secrete human erythropoietin in vivo. Hum. Gene Ther. *12*, 1527-1541.

Bennett,J.H., Joyner,C.J., Triffitt,J.T., and Owen,M.E. (1991). Adipocytic cells cultured from marrow have osteogenic potential. J. Cell Sci. *99 (Pt 1)*, 131-139.

Boyan,B.D. and Caplan,A.I. (1994). Endochondral Bone Formation: The Lineage Cascade. In Bone, CRC Press), pp. 1-46.

Boyne,P.J., Marx,R.E., Nevins,M., Triplett,G., Lazaro,E., Lilly,L.C., Alder,M., and Nummikoski,P. (1997). A feasibility study evaluating rhBMP-2/absorbable collagen sponge for maxillary sinus floor augmentation. Int. J. Periodontics. Restorative. Dent. *17*, 11-25.

Bruder,S.P., Fink,D.J., and Caplan,A.I. (1994). Mesenchymal stem cells in bone development, bone repair, and skeletal regeneration therapy. J. Cell Biochem. *56*, 283-294.

Bruder,S.P., Kraus,K.H., Goldberg,V.M., and Kadiyala,S. (1998). The effect of implants loaded with autologous mesenchymal stem cells on the healing of canine segmental bone defects. J. Bone Joint Surg. Am. *80*, 985-996.

Cadigan,K.M. and Nusse,R. (1997). Wnt signaling: a common theme in animal development. Genes Dev. *11*, 3286-3305.

Caplan,A.I. (1994). The mesengenic process. Clin. Plast. Surg. *21*, 429-435.

Caplan,A.I. and Bruder,S.P. (2001). Mesenchymal stem cells: building blocks for molecular medicine in the 21st century. Trends Mol. Med. *7*, 259-264.

Chaudhari,A., Ron,E., and Rethman,M.P. (1997). Recombinant human bone morphogenetic protein-2 stimulates differentiation in primary cultures of fetal rat calvarial osteoblasts. Mol. Cell Biochem. *167*, 31-39.

Dealy,C.N., Roth,A., Ferrari,D., Brown,A.M., and Kosher,R.A. (1993). Wnt-5a and Wnt-7a are expressed in the developing chick limb bud in a manner suggesting roles in pattern formation along the proximodistal and dorsoventral axes. Mech. Dev. *43*, 175-186.

Ducy,P., Schinke,T., and Karsenty,G. (2000). The osteoblast: a sophisticated fibroblast under central surveillance. Science *289*, 1501-1504.

Engstrand,T., Daluiski,A., Bahamonde,M.E., Melhus,H., and Lyons,K.M. (2000). Transient production of bone morphogenetic protein 2 by allogeneic transplanted transduced cells induces bone formation. Hum. Gene Ther. *11*, 205-211.

Fischer,L., Boland,G., and Tuan,R.S. (2002a). Wnt signaling during BMP-2 stimulation of mesenchymal chondrogenesis. J. Cell Biochem. *84*, 816-831.

Fischer,L., Boland,G., and Tuan,R.S. (2002b). Wnt-3A enhances bone morphogenetic protein-2-mediated chondrogenesis of murine C3H10T1/2 mesenchymal cells. J. Biol. Chem. *277*, 30870-30878.

Gaddy-Kurten,D., Coker,J.K., Abe,E., Jilka,R.L., and Manolagas,S.C. (2002). Inhibin suppresses and activin stimulates osteoblastogenesis and osteoclastogenesis in murine bone marrow cultures. Endocrinology *143*, 74-83.

Galotto,M., Campanile,G., Robino,G., Cancedda,F.D., Bianco,P., and Cancedda,R. (1994). Hypertrophic chondrocytes undergo further differentiation to osteoblast-like cells and participate in the initial bone formation in developing chick embryo. J. Bone Miner. Res. *9*, 1239-1249.

Gazit,D., Turgeman,G., Kelley,P., Wang,E., Jalenak,M., Zilberman,Y., and Moutsatsos,I. (1999). Engineered pluripotent mesenchymal cells integrate and differentiate in regenerating bone: a novel cell-mediated gene therapy. J. Gene Med. *1*, 121-133.

Gilbert,S.F. (2000). Developmental Biology. SINAUER ASSOCIATES, INC., Publishers).

Gong,Y., Slee,R.B., Fukai,N., Rawadi,G., Roman-Roman,S., Reginato,A.M., Wang,H., Cundy,T., Glorieux,F.H., Lev,D., Zacharin,M., Oexle,K., Marcelino,J., Suwairi,W., Heeger,S., Sabatakos,G., Apte,S., Adkins,W.N., Allgrove,J., Arslan-Kirchner,M., Batch,J.A., Beighton,P., Black,G.C., Boles,R.G., Boon,L.M., Borrone,C., Brunner,H.G., Carle,G.F., Dallapiccola,B., De Paepe,A., Floege,B., Halfhide,M.L., Hall,B., Hennekam,R.C., Hirose,T., Jans,A., Juppner,H., Kim,C.A., Keppler-Noreuil,K., Kohlschuetter,A., LaCombe,D., Lambert,M., Lemyre,E., Letteboer,T., Peltonen,L., Ramesar,R.S., Romanengo,M., Somer,H., Steichen-Gersdorf,E., Steinmann,B., Sullivan,B., Superti-Furga,A., Swoboda,W., van den Boogaard,M.J., Van Hul,W., Vikkula,M., Votruba,M., Zabel,B., Garcia,T., Baron,R., Olsen,B.R., and Warman,M.L. (2001). LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. Cell *107*, 513-523.

Groeneveld,E.H. and Burger,E.H. (2000). Bone morphogenetic proteins in human bone regeneration. Eur. J. Endocrinol. *142*, 9-21.

Hanada,K., Dennis,J.E., and Caplan,A.I. (1997). Stimulatory effects of basic fibroblast growth factor and bone morphogenetic protein-2 on osteogenic differentiation of rat bone marrow-derived mesenchymal stem cells. J. Bone Miner. Res. *12*, 1606-1614.

Haynesworth,S.E., Goshima,J., Goldberg,V.M., and Caplan,A.I. (1992). Characterization of cells with osteogenic potential from human marrow. Bone *13*, 81-88.

Horton,W.A. (1990). The Biology of Bone Growth . Growth Genet. Horm. *6*, 1-3.

Howell,T.H., Fiorellini,J., Jones,A., Alder,M., Nummikoski,P., Lazaro,M., Lilly,L., and Cochran,D. (1997). A feasibility study evaluating rhBMP-2/absorbable collagen sponge device for local alveolar ridge preservation or augmentation. Int. J. Periodontics. Restorative. Dent. *17*, 124-139.

Kahn,A.J. and Simmons,D.J. (1977). Chondrocyte-to-osteocyte transformation in grafts of perichondrium-free epiphyseal cartilage. Clin. Orthop. 299-304.

Katagiri,T., Yamaguchi,A., Ikeda,T., Yoshiki,S., Wozney,J.M., Rosen,V., Wang,E.A., Tanaka,H., Omura,S., and Suda,T. (1990). The non-osteogenic mouse pluripotent cell line, C3H10T1/2, is induced to differentiate into osteoblastic cells by recombinant human bone morphogenetic protein-2. Biochem. Biophys. Res. Commun. *172*, 295-299.

Kearns,A.E. and Demay,M.B. (2000). BMP-2 induces the expression of activin betaA and follistatin in vitro. J. Cell Biochem. *79*, 80-88.

Kengaku,M., Capdevila,J., Rodriguez-Esteban,C., De La,P.J., Johnson,R.L., Belmonte,J.C., and Tabin,C.J. (1998). Distinct WNT pathways regulating AER formation and dorsoventral polarity in the chick limb bud. Science *280*, 1274-1277.

Kenley,R., Marden,L., Turek,T., Jin,L., Ron,E., and Hollinger,J.O. (1994). Osseous regeneration in the rat calvarium using novel delivery systems for recombinant human bone morphogenetic protein-2 (rhBMP-2). J. Biomed. Mater. Res. *28*, 1139-1147.

Kim,K.J., Itoh,T., and Kotake,S. (1997). Effects of recombinant human bone morphogenetic protein-2 on human bone marrow cells cultured with various biomaterials. J. Biomed. Mater. Res. *35*, 279-285.

Kirker-Head,C.A., Gerhart,T.N., Schelling,S.H., Hennig,G.E., Wang,E., and Holtrop,M.E. (1995). Long-term healing of bone using recombinant human bone morphogenetic protein 2. Clin. Orthop. 222-230.

Kitagaki,J., Iwamoto,M., Liu,J.G., Tamamura,Y., Pacifci,M., and Enomoto-Iwamoto,M. (2003). Activation of beta-catenin-LEF/TCF signal pathway in chondrocytes stimulates ectopic endochondral ossification. Osteoarthritis. Cartilage. *11*, 36-43.

Knutsen,R., Wergedal,J.E., Sampath,T.K., Baylink,D.J., and Mohan,S. (1993). Osteogenic protein-1 stimulates proliferation and differentiation of human bone cells in vitro. Biochem. Biophys. Res. Commun. *194*, 1352-1358.

Lee,S.C., Shea,M., Battle,M.A., Kozitza,K., Ron,E., Turek,T., Schaub,R.G., and Hayes,W.C. (1994). Healing of large segmental defects in rat femurs is aided by RhBMP-2 in PLGA matrix. J. Biomed. Mater. Res. *28*, 1149-1156.

Lieberman,J.R., Le,L.Q., Wu,L., Finerman,G.A., Berk,A., Witte,O.N., and Stevenson,S. (1998). Regional gene therapy with a BMP-2-producing murine stromal cell line induces heterotopic and orthotopic bone formation in rodents. J. Orthop. Res. *16*, 330-339.

Liechty,K.W., MacKenzie,T.C., Shaaban,A.F., Radu,A., Moseley,A.M., Deans,R., Marshak,D.R., and Flake,A.W. (2000). Human mesenchymal stem cells engraft and demonstrate site-specific differentiation after in utero transplantation in sheep. Nat. Med. *6*, 1282-1286.

Lou,J., Xu,F., Merkel,K., and Manske,P. (1999). Gene therapy: adenovirus-mediated human bone morphogenetic protein-2 gene transfer induces mesenchymal progenitor cell proliferation and differentiation in vitro and bone formation in vivo. J. Orthop. Res. *17*, 43-50.

Mackay,A.M., Beck,S.C., Murphy,J.M., Barry,F.P., Chichester,C.O., and Pittenger,M.F. (1998). Chondrogenic differentiation of cultured human mesenchymal stem cells from marrow. Tissue Eng *4*, 415-428.

Maliakal,J.C., Asahina,I., Hauschka,P.V., and Sampath,T.K. (1994). Osteogenic protein-1 (BMP-7) inhibits cell proliferation and stimulates the expression of markers characteristic of osteoblast phenotype in rat osteosarcoma (17/2.8) cells. Growth Factors *11*, 227-234.

Mankani,M.H., Kuznetsov,S.A., Fowler,B., Kingman,A., and Robey,P.G. (2001). In vivo bone formation by human bone marrow stromal cells: effect of carrier particle size and shape. Biotechnol. Bioeng. *72*, 96-107.

Marden,L.J., Hollinger,J.O., Chaudhari,A., Turek,T., Schaub,R.G., and Ron,E. (1994). Recombinant human bone morphogenetic protein-2 is superior to demineralized bone matrix in repairing craniotomy defects in rats. J. Biomed. Mater. Res. *28*, 1127-1138.

Marks,S.C.Jr. and Odgren,P.R. (2002). **Structure and Development of the Skeleton**. In **Principles of Bone Biology 2 Volume Set**, J.P.Bilezikian, L.G.Raisz, and G.A.Rodan, eds. **Academic Press**), pp. 3-15.

Moon,R.T., Brown,J.D., and Torres,M. (1997). WNTs modulate cell fate and behavior during vertebrate development. Trends Genet *13*, 157-162.

Moutsatsos,I.K., Turgeman,G., Zhou,S., Kurkalli,B.G., Pelled,G., Tzur,L., Kelley,P., Stumm,N., Mi,S., Muller,R., Zilberman,Y., and Gazit,D. (2001). Exogenously regulated stem cell-mediated gene therapy for bone regeneration. Mol. Ther. *3*, 449-461.

Nolta,J.A., Hanley,M.B., and Kohn,D.B. (1994). Sustained human hematopoiesis in immunodeficient mice by cotransplantation of marrow stroma expressing human interleukin-3: analysis of gene transduction of long-lived progenitors. Blood *83*, 3041-3051.

Parr,B.A. and McMahon,A.P. (1995). Dorsalizing signal Wnt-7a required for normal polarity of D-V and A-P axes of mouse limb. Nature *374*, 350-353.

Pelled,G., G T, Aslan,H., Gazit,Z., and Gazit,D. (2002). Mesenchymal stem cells for bone gene therapy and tissue engineering. Curr. Pharm. Des *8*, 1917-1928.

Perrimon,N. and Bernfield,M. (2000). Specificities of heparan sulphate proteoglycans in developmental processes. Nature *404*, 725-728.

Pittenger,M.F., Mackay,A.M., Beck,S.C., Jaiswal,R.K., Douglas,R., Mosca,J.D., Moorman,M.A., Simonetti,D.W., Craig,S., and

Marshak,D.R. (1999). Multilineage potential of adult human mesenchymal stem cells. Science *284*, 143-147.

Rickard,D.J., Sullivan,T.A., Shenker,B.J., Leboy,P.S., and Kazhdan,I. (1994). Induction of rapid osteoblast differentiation in rat bone marrow stromal cell cultures by dexamethasone and BMP-2. Dev. Biol. *161*, 218-228.

Roach,H.I., Erenpreisa,J., and Aigner,T. (1995). Osteogenic differentiation of hypertrophic chondrocytes involves asymmetric cell divisions and apoptosis. J. Cell Biol. *131*, 483-494.

Sampath,T.K., Coughlin,J.E., Whetstone,R.M., Banach,D., Corbett,C., Ridge,R.J., Ozkaynak,E., Oppermann,H., and Rueger,D.C. (1990). Bovine osteogenic protein is composed of dimers of OP-1 and BMP-2A, two members of the transforming growth factor-beta superfamily. J. Biol. Chem. *265*, 13198-13205.

Sandhu,H.S., Kanim,L.E., Kabo,J.M., Toth,J.M., Zeegen,E.N., Liu,D., Delamarter,R.B., and Dawson,E.G. (1996). Effective doses of recombinant human bone morphogenetic protein-2 in experimental spinal fusion. Spine *21*, 2115-2122.

Teitelbaum,S.L. (2000). Bone resorption by osteoclasts. Science *289*, 1504-1508.

Thesingh,C.W., Groot,C.G., and Wassenaar,A.M. (1991). Transdifferentiation of hypertrophic chondrocytes into osteoblasts in murine fetal metatarsal bones, induced by co-cultured cerebrum. Bone Miner. *12*, 25-40.

Toriumi,D.M., Kotler,H.S., Luxenberg,D.P., Holtrop,M.E., and Wang,E.A. (1991). Mandibular reconstruction with a recombinant bone-inducing factor. Functional, histologic, and biomechanical evaluation. Arch. Otolaryngol. Head Neck Surg. *117*, 1101-1112.

Tremain,N., Korkko,J., Ibberson,D., Kopen,G.C., DiGirolamo,C., and Phinney,D.G. (2001). MicroSAGE analysis of 2,353 expressed genes in a single cell-derived colony of undifferentiated human mesenchymal stem cells reveals mRNAs of multiple cell lineages. Stem Cells *19*, 408-418.

Turgeman,G., Aslan,H., Gazit,Z., and Gazit,D. (2002a). Cell-mediated gene therapy for bone formation and regeneration. Curr. Opin. Mol. Ther. *4*, 390-394.

Turgeman,G., Pittman,D.D., Muller,R., Kurkalli,B.G., Zhou,S., Pelled,G., Peyser,A., Zilberman,Y., Moutsatsos,I.K., and Gazit,D. (2001). Engineered human mesenchymal stem cells: a novel platform for skeletal cell mediated gene therapy. J. Gene Med. *3*, 240-251.

Turgeman,G., Zilberman,Y., Zhou,S., Kelly,P., Moutsatsos,I.K., Kharode,Y.P., Borella,L.E., Bex,F.J., Komm,B.S., Bodine,P.V., and Gazit,D. (2002b). Systemically administered rhBMP-2 promotes MSC activity and reverses bone and cartilage loss in osteopenic mice. J. Cell Biochem. *86*, 461-474.

Urist,M.R. (1965). Bone: formation by autoinduction. Science *150*, 893-899.

Van de Putte,K.A. and Urist,M.R. (1965). Osteogenesis in the interior of intramuscular implants of decalcified bone matrix. Clin. Orthop. *43*, 257-270.

van den Bergh,J.P., ten Bruggenkate,C.M., Groeneveld,H.H., Burger,E.H., and Tuinzing,D.B. (2000). Recombinant human bone morphogenetic protein-7 in maxillary sinus floor elevation surgery in 3 patients compared to autogenous bone grafts. A clinical pilot study. J. Clin. Periodontol. *27*, 627-636.

Volek-Smith,H. and Urist,M.R. (1996). Recombinant human bone morphogenetic protein (rhBMP) induced heterotopic bone development in vivo and in vitro. Proc. Soc. Exp. Biol. Med. *211*, 265-272.

Wang,E.A., Israel,D.I., Kelly,S., and Luxenberg,D.P. (1993). Bone morphogenetic protein-2 causes commitment and differentiation in C3H10T1/2 and 3T3 cells. Growth Factors *9*, 57-71.

Wang,E.A., Rosen,V., Cordes,P., Hewick,R.M., Kriz,M.J., Luxenberg,D.P., Sibley,B.S., and Wozney,J.M. (1988). Purification and characterization of other distinct bone-inducing factors. Proc. Natl. Acad. Sci. U. S. A *85*, 9484-9488.

Winnier,G., Blessing,M., Labosky,P.A., and Hogan,B.L. (1995). Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. Genes Dev. *9*, 2105-2116.

Wozney,J.M., Rosen,V., Byrne,M., Celeste,A.J., Moutsatsos,I., and Wang,E.A. (1990). Growth factors influencing bone development. J. Cell Sci. Suppl *13*, 149-156.

Wozney,J.M., Rosen,V., Celeste,A.J., Mitsock,L.M., Whitters,M.J., Kriz,R.W., Hewick,R.M., and Wang,E.A. (1988). Novel regulators of bone formation: molecular clones and activities. Science *242*, 1528-1534.

Yamaguchi,A., Ishizuya,T., Kintou,N., Wada,Y., Katagiri,T., Wozney,J.M., Rosen,V., and Yoshiki,S. (1996). Effects of BMP-2, BMP-4, and BMP-6 on osteoblastic differentiation of bone marrow-derived stromal cell lines, ST2 and MC3T3-G2/PA6. Biochem. Biophys. Res. Commun. *220*, 366-371.

Yamaguchi,A., Katagiri,T., Ikeda,T., Wozney,J.M., Rosen,V., Wang,E.A., Kahn,A.J., Suda,T., and Yoshiki,S. (1991). Recombinant human bone morphogenetic protein-2 stimulates osteoblastic maturation and inhibits myogenic differentiation in vitro. J. Cell Biol. *113*, 681-687.

Yasko,A.W., Lane,J.M., Fellinger,E.J., Rosen,V., Wozney,J.M., and Wang,E.A. (1992). The healing of segmental bone defects, induced by recombinant human bone morphogenetic protein (rhBMP-2). A radiographic, histological, and biomechanical study in rats. J. Bone Joint Surg. Am. *74*, 659-670.

Yoo,J.U., Barthel,T.S., Nishimura,K., Solchaga,L., Caplan,A.I., Goldberg,V.M., and Johnstone,B. (1998). The chondrogenic potential of human bone-marrow-derived mesenchymal progenitor cells. J. Bone Joint Surg. Am. *80*, 1745-1757.

Young,R.G., Butler,D.L., Weber,W., Caplan,A.I., Gordon,S.L., and Fink,D.J. (1998). Use of mesenchymal stem cells in a collagen matrix for Achilles tendon repair. J. Orthop. Res. *16*, 406-413.

Zegzula,H.D., Buck,D.C., Brekke,J., Wozney,J.M., and Hollinger,J.O. (1997). Bone formation with use of rhBMP-2 (recombinant human bone morphogenetic protein-2). J. Bone Joint Surg. Am. *79*, 1778-1790.

Zhang,H. and Bradley,A. (1996). Mice deficient for BMP2 are nonviable and have defects in amnion/chorion and cardiac development. Development *122*, 2977-2986.

Chapter 4

Searching for Mutant p53 Target Genes

Introduction

Tumor Suppressor p53

The tumor suppressor protein p53 plays a critical role in controlling cell life and death. P53 functions as a transcription factor and is one of the main players in a large network of messenger proteins and effectors that are involved with the G1 and G2 DNA damage checkpoints.

P53 is activated in stressed and damaged cells, which have greater tendency to become cancerous. Several types of stress can activate p53, including DNA damage and oncogene activation, hypoxia, depletion of the cell's nucleotide pool, or defects in DNA methylation (Soussi and Beroud, 2001). Once activated, p53 activates the transcription of numerous target genes including genes that block cell division, genes that cause apoptosis, genes that regulate DNA repair, and genes that inhibit blood vessel formation. This transcriptional activity is critical to p53's role as a tumor suppressor (Bullock and Fersht, 2001; Vousden, 2000).

In normal cells, p53 is found at very low concentrations, its levels are regulated by two interactive feedback loops (Jin and Levine, 2001). One loop involves p53 and mdm2, an ubiquitin ligase that negatively regulates p53 concentrations by targeting p53 for degradation in the proteasome (Soussi and Beroud, 2001)(Figure 1). P53 positively regulates mdm2 by activating its transcription.

The other loop involves p14ARF and p53 in which p14ARF interacts with mdm2 , inhibiting the mdm2 mediated degradation of p53. p53, in turn, suppresses ARF transcription (Jin and Levine, 2001).

Various cellular stress signals activate p53 through upstream mediators, mostly by post translational modifications. These modifications stabilize p53 by interfering with the mdm2-p53 interaction and also serve to localize p53 to the nucleus and regulate its transcriptional activity.

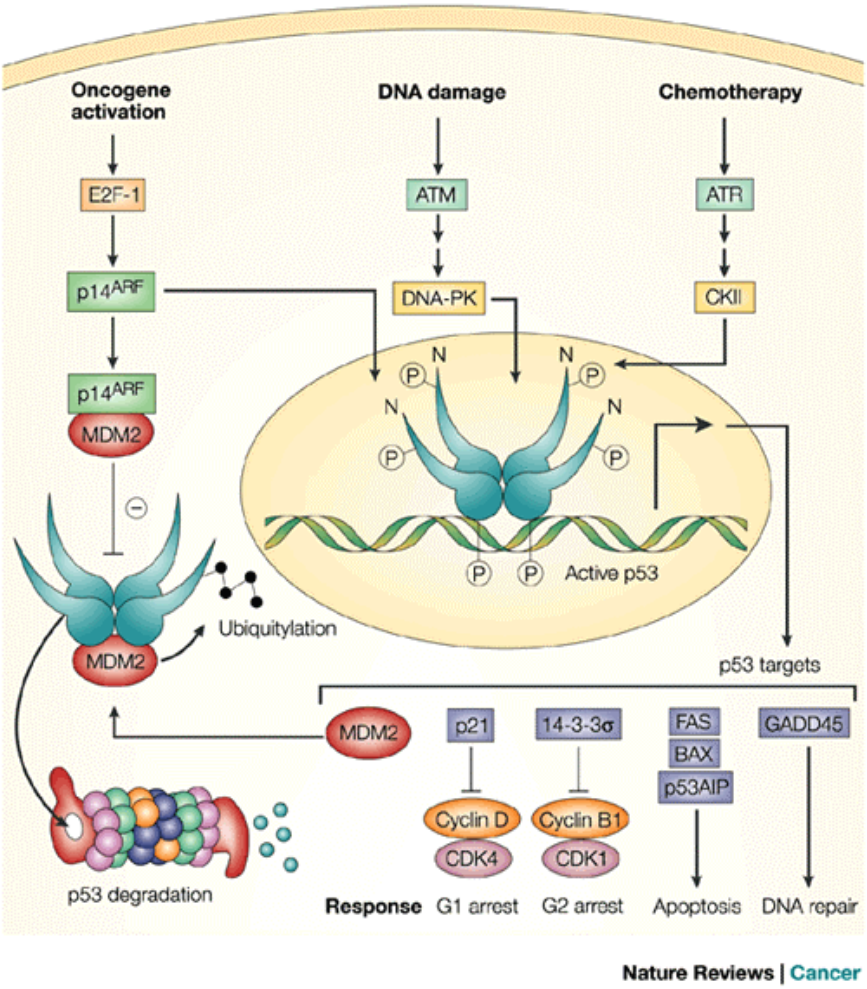

Figure 1 | p53 Response to Stress. Cellular stress such as Oncogene activation, DNA damage and chemotherapy activate signal-transduction pathways that lead to post-translational modifications of p53 and/or mdm2, thus stabilizing p53 and increasing its sequence specific DNA binding. As a result, trancription of p53 target genes is activated. These target genes include p21 leading to G1 arrest, 14-3-3sigma, leading to G2 arrest, apoptosis promoting genes (FAS,BAX) ,DNA repair genes (GADD45), and mdm2. (adapted from Bullock and Fersht, 2001).

P53 Structural Motifs

P53 is a 393 amino-acid protein and has the usual hallmarks of a transcription factor (Figure 2a) (Bullock and Fersht, 2001; Levine, 1997; Ahn

and Prives, 2001), including an amino terminal transactivation domain (amino acids 1-~50) that interacts with the general transcription machinery such as RNA polimerase and other factors, a core DNA binding domain (amino acids 102-292), a carboxy-terminal tetramerization domain (amino acids 324-355, Figure 4) (p53 works as a tetramer) and regulatory domains (amino acids 363-393).

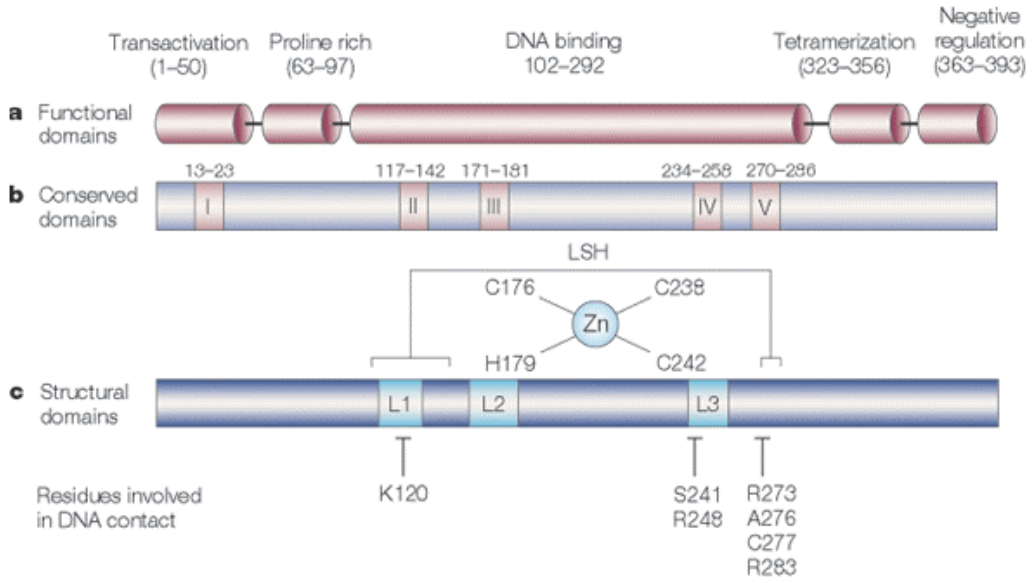

Figure 2 | **Schematic view of p53. a**. p53s functional domains. **b**. p53s conserved regions. **c**. p53 structural domains, L- Loop, LSH-Loop Sheet Helix. Residues involved in DNA contact are shown and residues involved in Zn binding (adapted from Soussi and Beroud, 2001)**.**

The p53 core DNA binding domain is responsible for the sequence specific binding of p53 to its consensus sequences. This domain independently folds into four stranded and five stranded anti-parallel beta-sheets; a loop-sheet-helix motif is packed against one of the ends of this beta sandwich (Cho et al., 1994). Additionally, there are two large loops that are held together in part by a zinc atom. The loop-sheet-helix motif and one of the two large loops are used to bind DNA (Figure 2c, 3).

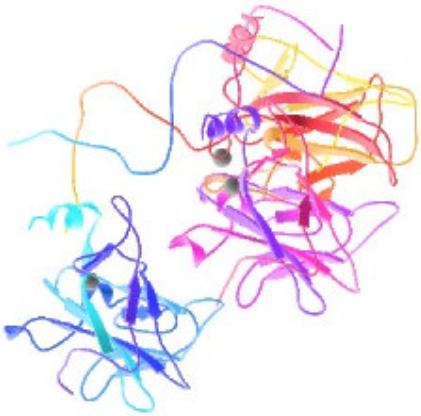

Figure 3| p53 DNA binding core domain bound to DNA (amino acids 102-293). Zinc atoms are shown in gray Taken from the protein data bank.

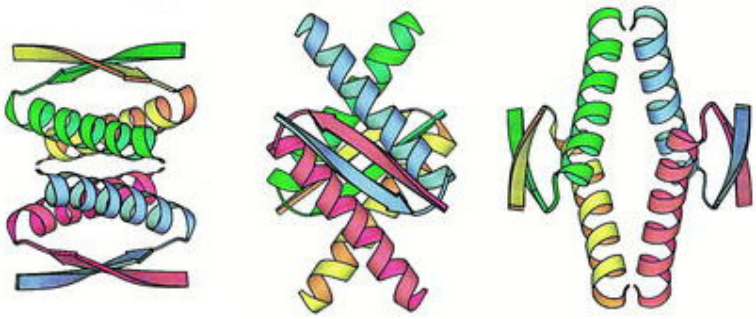

Figure 4| Ribbon diagram of the p53 tetramerisation domain residues 326 to 356 (four subunits). The tetramerisation domain has 222 point symmetry. The diagrams are projections along the three 2-fold axes. Each of the four different subunits is colored in a differently: red, blue, green or yellow.

Mutant p53

Mutations of p53 are associated with about 50% of all human cancers, and probably confer some selective advantage in carcinogenesis. Most p53 mutations (~75%) are single misssense mutation (Bullock and Fersht, 2001). More than 10,000 cancer related p53 mutations have been discovered; 95% of them are in the p53 core domain. These mutations can be divided into two classes: (a) mutation in amino acids that are in direct contact with the DNA: There are two mutation "hotspots" R248 (Loop 3) or R273 (Sheet 10 in the LoopSheetHelix). (b) mutations in amino acids that stabilize the structure of the core domain: there are four "hotspots" for mutation, R175

(Loop2), G245 (Loop3), R249 (Loop3) and R282 (Helix 2 in LoopSheetHelix) (Bullock and Fersht, 2001)(Figure 5). At normal body temperature all these "hotspot" mutants fail to bind to the known p53 consensus sequence. Thus, p53 does not transactivate its' usual set of genes, including mdm2. As a result, the mutant p53 half-life is increased and mutated p53 accumulates at high levels in the cell.

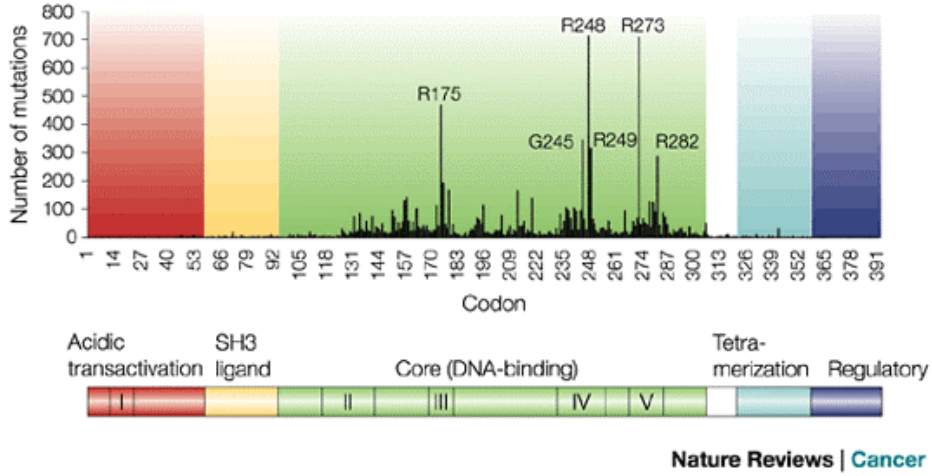

Figure 5 | p53 Conserved Regions. p53s regions of greatest sequence conservation are shown; four map to the core domain. A histogram of p53 missense mutations shows that 95% of mutations occur in the core domain; six labelled residues are hot spots for mutation (adapted from Bullock and Fersht, 2001).

Cancers with p53 mutation are often aggressive and more resistant to therapy (Bullock and Fersht, 2001). This finding can be explained by "loss of function": the cells may lack wild type p53 (in many tumor cells that have a mutation in one allele there is a deletion of the p53 gene in the other allele), or the mutant p53 exerts a dominant negative effect on the w.t. p53 (the mutant p53 forms non-functional oligamers with w.t. p53) (van Oijen and Slootweg, 2000; Bullock and Fersht, 2001). However, several studies have shown that mutant p53 might also confer a protective effect of its own (Blandino et al., 1999; van Oijen and Slootweg, 2000). This hypothesis, of "gain of function" by mutant p53, is supported by the finding that over expression of some forms of mutant p53 enhances the resistance to anti-cancer agents compared to p53 null cells (Blandino et al., 1999). Thus, these

mutants not only interfere with p53 dependent apoptosis, through a dominant negative mechanism, but also with p53 independent apoptosis. In addition to that, core domain p53 mutants were shown to increase mutation frequency (Iwamoto et al., 1996), block differentiation (Shaulsky et al., 1991) and increase metastatic potential (Hsiao et al., 1994; Crook and Vousden, 1992). The precise mechanisms of this oncogenic gain of function are unknown, but they may include transcriptional transactivation.
Further supporting this is the finding that the integrity of the N-terminal transactivation domain of p53 is required for mutant p53 interference with drug induced apoptosis (Matas et al., 2001).

Our hypothesis is that mutant p53 functions as a transcription factor that specifically transactivates target genes in the process of the malignant transformation. Our goal is to identify such genes using DNA chips.

The Experimental System

In order to test this hypothesis a series of DNA chip experiments were preformed, testing the gene expression levels of cells containing mutant p53 and the gene expression levels of p53 null cells.
The experiments were designed and preformed by Lilach Weis at Prof. Varda Rotters' lab.
Lilach used the human lung cancer cell line H1299 which lacks endogenous p53.
Two p53 mutant forms were chosen:
1. Contact mutation: Cells with p53 gene mutated at position 273 Arg → His (R273H).
2. Structural mutation: Cells with p53 gene mutated at position 175 Arg → His (R175H).
Both mutations show gain of function characteristics (8,9-proposal) and both were shown to induce transactivation of specific genes (10-14-proposal).

Introduction of Mutant p53 into p53 null cells

The mutant forms were introduced to the cells in two different ways: stable transfection and infection.

Transfection: Plasmids containing the R175H p53 mutant, the R273H p53 mutant, or an empty plasmid (for control) were introduced into cells. Neo resistant clones were selected using G418. Out of the resistant R175H mutant clones, 10 were found to express high levels of mutant p53, out of the resistant R273H clones 7 were found to express the mutant protein at high levels.

Infection: Cells were transfected with an ecotropic receptor, to enable the use of the murine retroviral vector. Afterwards, the cells were infected with virus, carrying either the R175H p53 mutant, the R273H p53 mutant or the empty plasmid. Cells were selected using puromycin. 90% of the cells passing selection were found to express mutant p53 at high levels.

RNA extraction

RNA was extracted from each of the 10 chosen R175H mutant transfected clones and pooled. The same was done for the 7 clones expressing the R273H mutant. As control, RNA was extracted from cells transfected with an empty vector.

For the transfected cells, this procedure was repeated 3 times, each time clones were grown from frozen stocks.

For the infection this procedure was repeated twice, each time H1299 were infected.

RNA samples were labeled.

Hybridization and Scanning

RNA samples were hybridized with DNA Affymetrix human GeneChip 6500. The chips were scanned at the DNA chip center and processed using the Affymetrix Genechip 3.3 software.

The results were then brought to us for further analysis.

Methods

Dataset

Our initial dataset consisted of 15 experiments, composed of 5 control experiments, 5 R175H mutant p53 experiments, and 5 R273H mutant p53 experiments. Each quintet is composed of 3 transfections and 2 infections. Each chip measured the expression level of 7,129 genes. So our dataset was a matrix of size 15*7129.

At first the analysis was done using the full size matrix. Later on, the reliability of some of the experiments (all of the infections+ the first series of transfections) was questioned and they were taken out of the analysis, leaving a 6*7129 matrix.

Pre-Processing

The expression data was organized in a matrix of 7,100 rows (genes on the chip) and $n_s=15$ and $n_s=6$ columns (experiments), for the full dataset and the reduced dataset, respectively. First, we removed non-human Affymetrix controls. The remaining data was preprocessed and filtered as described in the methods section using a threshold value of T=10 for the thresholding operation and $\sigma_T = 0.7$ for the filtering operation. The number of genes that passed these procedures are 1469 and 1263, for the full and reduced datasets, respectively.

Supervised Analysis

We wanted to find genes whose expression levels change between the different groups of samples: cells containing empty plasmid, cells containing p53 mutated in R273H, and cells containing p53 mutated in R175H. In order to find such known classes in data, a good strategy to use is that of supervised statistical analysis. The advantages of such an analysis are: (i)

focus on the relevant biological process that is of interest and (ii) easy determination of the statistical significance of our results.

We use hypothesis testing in order to identify genes that are differentially expressed between the different populations of cells.

We preformed a series of tests comparing two groups of samples at a time:

1. Empty Cells Vs. R175H mutant cells
2. Empty Cells Vs. R273H mutant cells
3. R175H mutant cells Vs. R273H mutant cells
4. Empty Vs. R175H mutant cells+R273H mutant cells

The Statistical Tests

Two statistics were used, gene by gene, for this analysis: the *t-value* statistic, which is the standardized difference between the means of two distributions, and, the d-value, the distance (see below) between the means of two distributions. The t-value was used at the first stage of the analysis, when all 15 samples were analyzed. For the analysis of the smaller data set (only two replicates for each type of cells), the d-value was used because it is not possible to get a good estimation of the standard deviation from only two samples.

We used these statistics in order to test the null hypothesis that the expression values of group A are taken from the same distribution as the expression values of group B.

The test goes as follows: A t-value (or d-value) is calculated for each gene. For each calculated t-value a p-value is obtained. The p-value is the probability of getting such a t-value, or higher, under our null hypothesis. If our p-value is small enough we can reject the null hypothesis and accept the alternative hypothesis: the expression values of group A are taken from a different distribution than the expression values of group B. Thus, this particular gene may be related to the biological process of interest.

Calculating the t-value/d-value: Let e_{gs} be the log2 of the expression level of gene *g* in sample *s*. $\overline{X}_A$ is the mean of the e_{gs} values that belong to samples of group A, σ_A is their standard deviation, and n_A is the number of samples in group A. $\overline{X}_B$ is the mean of e_{gs} values that belong to samples of group B, σ_B is their standard deviation, and n_B is the number of samples in group B.

The t-value is calculated as follows:

$$t = \left|\overline{X}_A - \overline{X}_B\right| \Big/ \sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}$$

The d-value:

$$d = \left|\overline{X}_A - \overline{X}_B\right|$$

Usually, after obtaining a t-value, one would go to the standard tables and obtain the corresponding P-value. These tables are calculated under the assumption that the data are taken from a normal distribution. In our case we cannot make this assumption; so we use a *numerical* method, based on enumerating all permutations, in order to estimate the distribution of t-values and thus, the corresponding P-value.

Obtaining the P-value: In order to obtain the P-value, all the possibilities for partitioning the $n_A + n_B$ samples into group A' with n_A samples and group B' with the remaining n_B samples were considered and for each such partitioning a t-value was calculated as before. The resulting distribution of t – values, ρ(t) is used as our estimate of the t-value distribution in the case when the null hypothesis holds.
Now, for each previously obtained t-value, t_v , we calculate a P-value, the probability to obtain a value $t \geq t_v$ drawn from the distribution ρ(t):

$$P = \int_{t_v}^{\infty} dt\rho(t)$$

Thus, for each gene we have a P-value, the probability that we erred in rejecting the null hypothesis. The lower the P-value, the more confidence we have that this gene separates between the two groups A and B. The same was done for the d-value statistic: for each permutation a d was calculated. There is a total of 15 such possibilities (6!/2!4!), thus a total of 15 different d-values. Now for each previously obtained d-value, d_v we calculate a P-value, the probability to generate a value $d \geq d_v$ drawn from the distribution ρ(d):

$$P = \sum_{d \geq dv} \frac{1}{15}$$

Since we only have 15 permutations our lowest p-value will be 1/15=0.0667.

Unsupervised Analysis

Before the clustering procedure each gene's expression levels were centered and normalized (see methods section).

The SPC clustering algorithm was used in order to cluster the genes according to their expression levels in the 6 reliable samples. Trying to do so, we found that most of the genes that passed our filtering, did not behave consistently throughout the replicate experiments.

We decided to filter the genes in away that will leave us with genes that act consistently throughout "identical" sets of experiments.

The idea was to select genes that most of the variance in their expression level arises from the different conditions and not from the repeats of experiments.

The 6 samples were divided into the three experimental groups of two samples each: Empty, mutant p53 at position 273, and mutant p53 at position 175. Only genes that satisfied the condition $SS_{bg} > 2/3 SS_T$, where

SS_{bg} denotes the Sum of Squares between the groups and SS_T total Sum of Squares, and satisfied std>0.5, were taken.

Calculation of Sum of Squares for each gene:

M_g -Mean of group g of samples, M_T -Total mean, G is the set of sample groups, N_g -The number of elements in the group g, I is the set of elements, "X_i" is the ith element.

$$SS_{bg} = \sum_{g \in G} (M_g - M_T)^2 N_g$$

$$SS_T = \sum_{i \in I} (X_i - M_T)^2$$

Results and Discussion

Supervised analysis

Full Data Set (15 experiments): The number of genes passing our filtering criteria of std>0.7 were 1469. Out of 1469 genes with randomly selected expression values (from the same distribution) one would expect to get 73 with a p-value of 0.05 or less.

We actually found the following number of genes in the four comparisons made: For R175H vs Empty- 79 genes; for R273H vs Empty- 49 genes; for R175H vs R273H- 40 genes; and Empty versus R175H+R273H yielded 80 genes.

Thus, the number of genes we got is less then or about equal to the expected number for random data, and we cannot draw any conclusions about these genes.

Reduced Data Set (6 experiments): The number of genes that passed our filtering criteria of std>0.7 was 1263. Of them, a total of 198 genes had a p-value of 0.0667 (see Appendix D tables 1, 2) in the Empty versus R175H+R273H comparison. One would expect that from 1263 random genes there should be about 88 with such a p-value or less; thus, we have more genes that differentiate the Empty group from the joined R175H+R273H group, than we would expect from random data.

A larger set of samples is needed in order to refine our p-value resolution and to pinpoint a more specific subset of these genes.

Unsupervised Analysis

372 genes passed our filtering criteria (SSbg>2/3SSt, std>0.5); each gene's expression levels were centered and normalized. Interestingly, almost all the genes passing our filtering have different expression levels between the

empty cells and the mutated cells, but there are almost no genes having different expression levels between the two mutations.

Clustering of the data revealed two major clusters (Figure 4,5): G2 containing 180 genes (see Appendix D table 3), and G3 containing 111 genes (see Appendix D table 4). The clusters exhibit opposite behaviors: genes in the G2 cluster have a high expression level in the Empty cells relatively to both mutants (Figure 6a), and G3 genes have a low expression level in the Empty cells in comparison to the mutants (Figure 6b). The filtering procedure we preformed selects for genes which act consistently within each experimental group (Empty, R175H, and R273H) but varies among groups. This filtering is symetrical and except for the previous requirments it does not prefer one type of gene behavior over of the other. Thus, we would expect to get six more or less equally sized clusters of genes. The six possibilities are summarized in the table:

Empty	**R175H**	**R273H**
+	–	–
-	+	+
+	+	-
–	–	+
-	+	–
+	-	+

+ means expression values above average

- means expression values below average

Our set of genes contained the first two behaviors almost exclusively. The p-value for such a gene behavior distribution is below 0.001, calculated by the χ^2 test (Lowry, 2003). This is a very low p-value, indicating that this behavior is very far from that of random data.

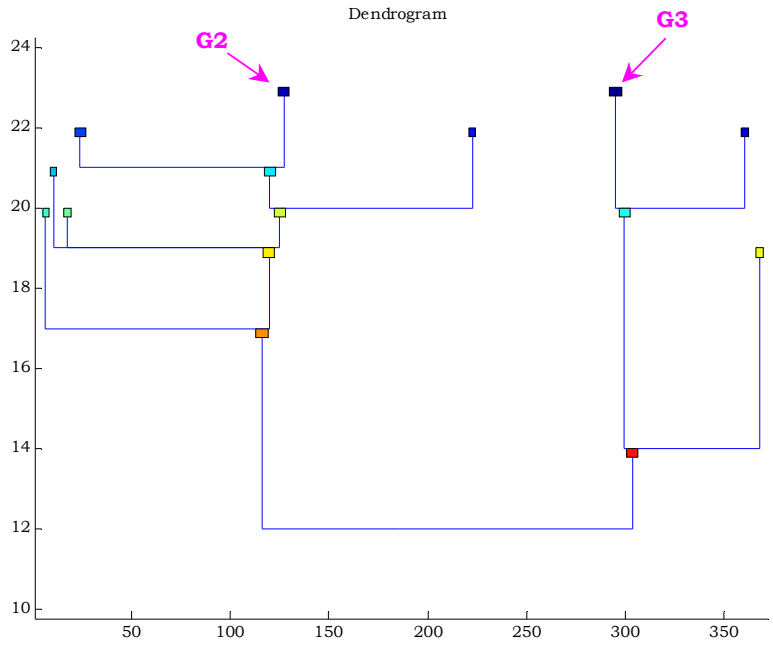

Figure 4 | Dendrogram of the Gene Clusters. G2 and G3 are indicated.

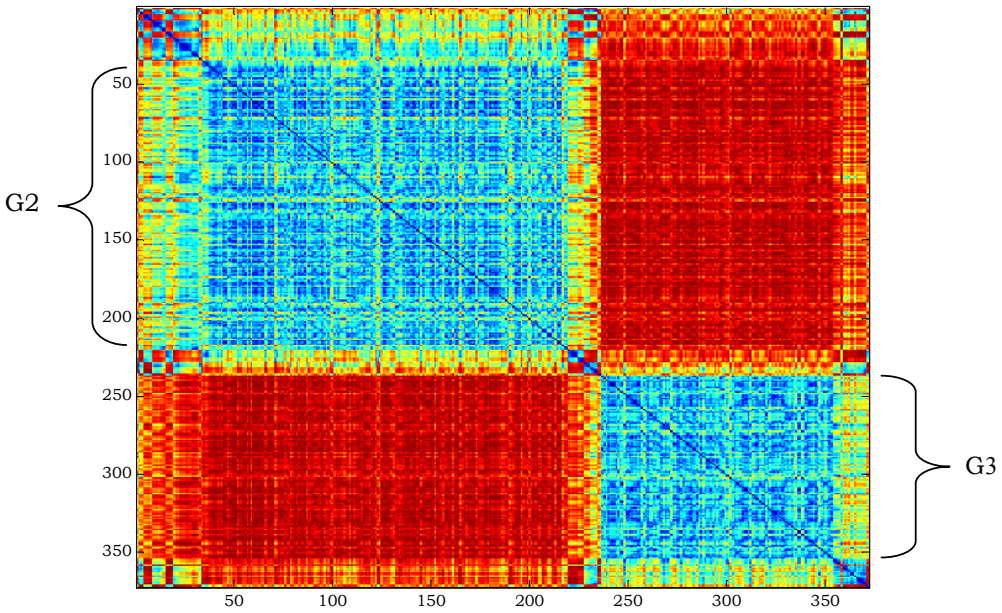

Figure 5 | Distance matrix of the genes ordered according to the clustering results. The Eucleadian distance is used. Blue-very close genes, Red-very distant genes.

Comparing with the results of the supervised analysis: 85 of the genes that were found to be upregulated in the p53 null cells in the statistical analysis passed our filtering criteria and all 85 of them were in cluster G2. 51 of the genes that were found to be downregulated passed our filtering and 50 of them were present in G3.

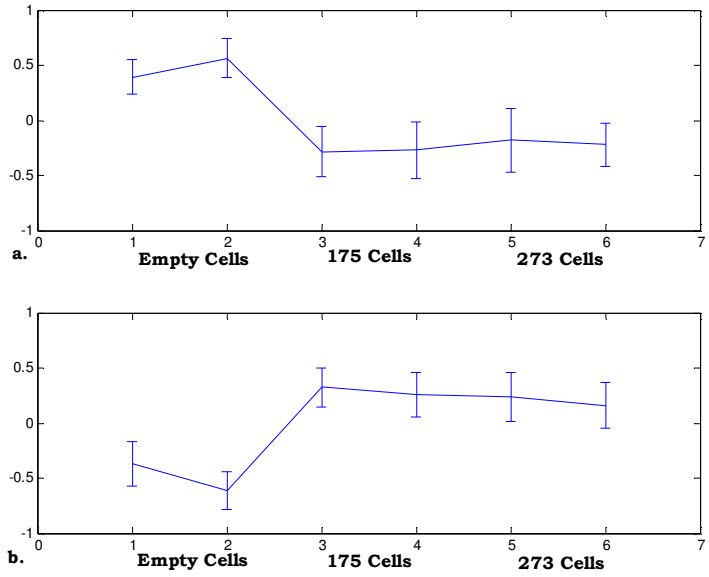

Figure 6 | **a**| profiles of G2 genes. **b**| profiles of G3 genes.

Relaxing the filtering criteria : This clustering procedure was repeated again, on a larger set of genes passing the SSbg>2/3SSt criteria, but this time, no lower threshold was set for their overall standard deviation. This filtering procedure left us with 1372 genes. Clustering of these genes separated the data in a similar way as before, into two main clusters: G2b with 567 genes, and G3b with 480 genes. These two clusters exhibit the same behaviors as G2 and G3. Indeed, 164 genes of the 180 genes of G2 (91%) are in G2b, and 107 genes out of the 111 genes of G3 (96%) are in G3b. The p-value for such a distribution of genes is very close to zero and is even lower than the one obtained before. This test indicates that the results are robust against changing the (arbitrary) value of the filter used.

The fact that the data separates into these two disticnt clusters, indicates that insertion of mutant p53 into null cells causes a significant change in the transcription of many genes. However, it seems that there is no significant difference between the two mutations.

Conclusions

Using all 15 experiments, our analysis did not find statistically significant results. This is in agreement with the fact that 9 out of these 15 experiments were found to be not reliable.

The analysis of the remaining 6 experiments revealed a statistically significant number of genes (both in the supervised and unsupervised analyses) having different expression levels between the Empty plasmid cells and the Mutant p53 population of cells. Still, the small number of samples does not allow for a more clear-cut statement regarding each of these genes; since the p-values are quantized, and the smallest value being 1/15. At this p-value about 90 genes out of the 1263 genes are expected to be falsely discovered. We found 200 genes with this p-value so each gene has a 9/20=45% chance of being false.

The big difference in gene expression observed here is between the empty plasmid containing cells (control) and the cells containing one of the p53 mutations. There does not seem to be a big difference in the gene behaviors between the two mutations. This may lead us to the conclusion that there is mutant p53 “gain of function” with respect to p53 null cells and that this gain of function is similar in both mutants. Perhaps it is a good idea to perform these chip experiments also with cells containing w.t. p53, to rule out the possibility that these genes change their level also in the presence of w.t. p53 and therefore are not associated with mutant p53 gain of function.

Reference List

Ahn,J. and Prives,C. (2001). The C-terminus of p53: the more you learn the less you know. Nat. Struct. Biol. *8*, 730-732.

Blandino,G., Levine,A.J., and Oren,M. (1999). Mutant p53 gain of function: differential effects of different p53 mutants on resistance of cultured cells to chemotherapy. Oncogene *18*, 477-485.

Bullock,A.N. and Fersht,A.R. (2001). Rescuing the function of mutant p53. Nat. Rev. Cancer *1*, 68-76.

Cho,Y., Gorina,S., Jeffrey,P.D., and Pavletich,N.P. (1994). Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science *265*, 346-355.

Crook,T. and Vousden,K.H. (1992). Properties of p53 mutations detected in primary and secondary cervical cancers suggest mechanisms of metastasis and involvement of environmental carcinogens. EMBO J. *11*, 3935-3940.

Hsiao,M., Low,J., Dorn,E., Ku,D., Pattengale,P., Yeargin,J., and Haas,M. (1994). Gain-of-function mutations of the p53 gene induce lymphohematopoietic metastatic potential and tissue invasiveness. Am. J. Pathol. *145*, 702-714.

Iwamoto,K.S., Mizuno,T., Ito,T., Tsuyama,N., Kyoizumi,S., and Seyama,T. (1996). Gain-of-function p53 mutations enhance alteration of the T-cell receptor following X-irradiation, independently of the cell cycle and cell survival. Cancer Res. *56*, 3862-3865.

Jin,S. and Levine,A.J. (2001). The p53 functional circuit. J. Cell Sci. *114*, 4139-4140.

Levine,A.J. (1997). p53, the cellular gatekeeper for growth and division. Cell *88*, 323-331.

Lowry, R. Concepts and Applications of Inferential Statistics. Vassar College . 2003.
Ref Type: Electronic Citation

Matas,D., Sigal,A., Stambolsky,P., Milyavsky,M., Weisz,L., Schwartz,D., Goldfinger,N., and Rotter,V. (2001). Integrity of the N-terminal transcription domain of p53 is required for mutant p53 interference with drug-induced apoptosis. EMBO J. *20*, 4163-4172.

Shaulsky,G., Goldfinger,N., and Rotter,V. (1991). Alterations in tumor development in vivo mediated by expression of wild type or mutant p53 proteins. Cancer Res. *51*, 5232-5237.

Soussi,T. and Beroud,C. (2001). Assessing TP53 status in human tumours to evaluate clinical outcome. Nat. Rev. Cancer *1*, 233-240.

van Oijen,M.G. and Slootweg,P.J. (2000). Gain-of-function mutations in the tumor suppressor gene p53. Clin. Cancer Res. *6*, 2138-2145.

Vousden,K.H. (2000). p53: death star. Cell *103*, 691-694.

Chapter 5

Methods

DNA Microarrays

The phenotype of cells and tissues is determined by the genes that are expressed at a given time. These genes are transcribed into mRNA and then translated into proteins. DNA microarrays measure the concentration of mRNA present, and hence the transcriptional activity of the genes. Although the basic technique for the detection of the mRNA levels is not new, the novelty in the DNA arrays is their ability to do so for thousands of genes simultaneously (Lockhart et al., 1996). This makes them a powerful tool for gene research.

The "gene chip" is actually an affinity matrix made of silicon or a glass slide or some other surface to which DNA molecules ("probes") are attached at very specific places. This chip works by hybridization of labeled RNA or DNA ("targets"), taken from a sample, to their specific complementary probe. The probes can be either large probes up to 5000bps long (cDNA, cDNA microarrays), or short up to 25bps long (olignucleotides, olignucleotide microarrays) (Lockhart and Winzeler, 2000).

We will focus on oligonucleotide microarrays.

Oligonucleotide Microarrays

The major manufacturer of oligonucleotide microarrays is Affymetrix. On Affymetrix chips each gene is represented by a probe set, consisting of a series of paired probe cells (usually 20) (Figure 1). Each probe cell contains millions of copies of a unique single strand DNA probe, 25 oligonucleotide long. Probes are tiled in probe pairs as a Perfect Match (PM) and a Mismatch (MM). The sequence for PM and MM are the same, except for a change of a single base in the middle of

the MM probe sequence. The MM is a measure for the background level of non specific hybridization (Lipshutz et al., 1999).

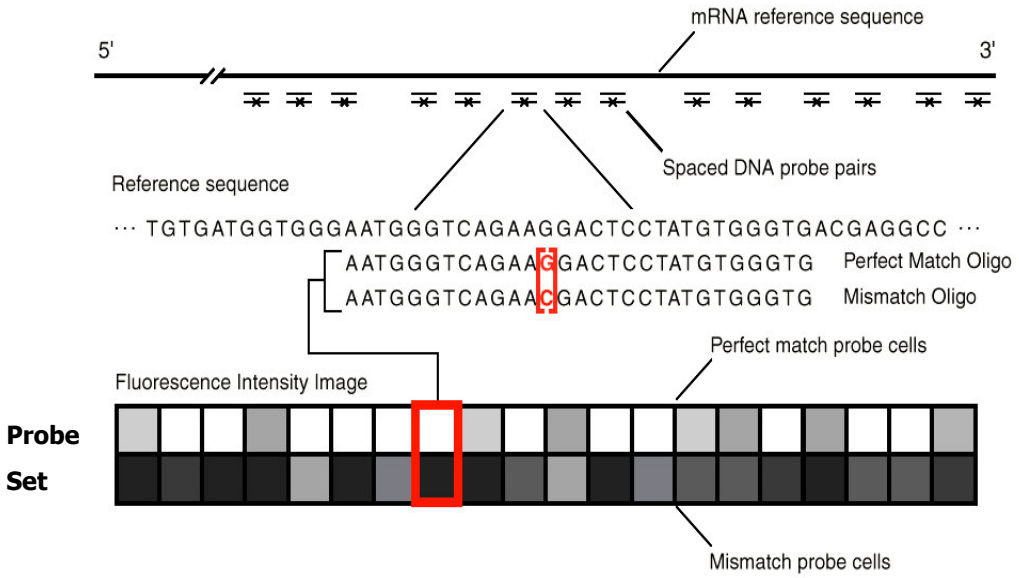

Figure 1 | Oligonecleotide Probes and the Probe Set. Probes are 25 base long oligonucleotide sequences chosen from the RNA reference sequence. The probe set contains 20 pairs of PM and MM probe cells. Each probe cell contains millions of copies of the cell-specific oligonecleotide probe.

The oligonucleotide probes are synthesized in situ on the surface of the chip using a sequence of light-directed coupling reaction (see figure 2). Using this process, termed photolithography, the glass is first covered with protecting groups that prevent DNA deposition. A mask is placed on the glass with gaps corresponding to the sites of nucleotide deposition. Then laser beams are shone onto the holes where synthesis is to begin. The light takes off the protecting groups. Then the glass is bathed in the first nucleotide to be attached. Each nucleotide carries its own protection group, which can be taken off for the second round of nucleotide attachments. Hence, by sequential application of the appropriate masks and bathing sequences, arrays of different nucleotides can be built up.

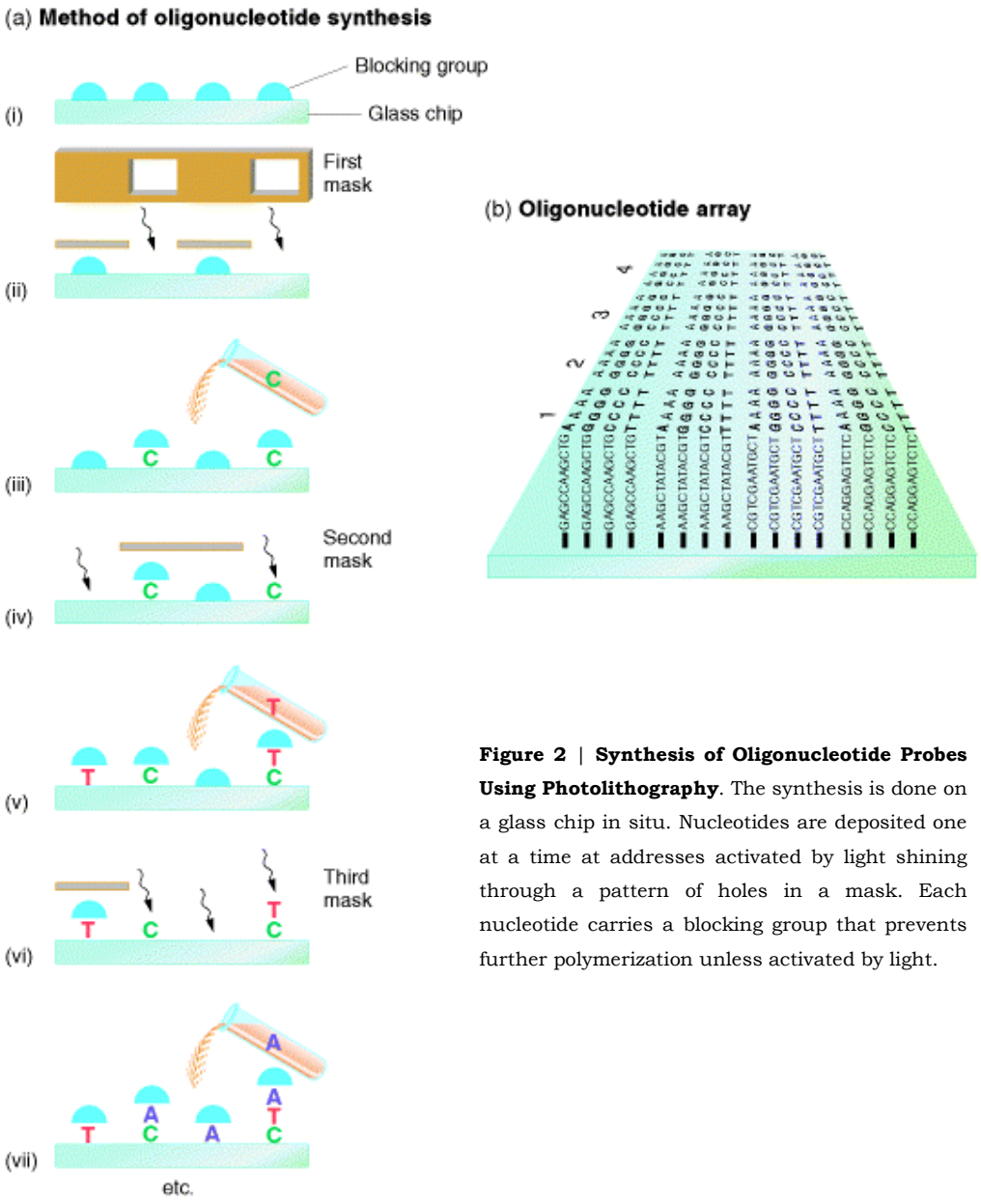

Figure 2 | Synthesis of Oligonucleotide Probes Using Photolithography. The synthesis is done on a glass chip in situ. Nucleotides are deposited one at a time at addresses activated by light shining through a pattern of holes in a mask. Each nucleotide carries a blocking group that prevents further polymerization unless activated by light.

Target Preparation and Hybridization

mRNA is extracted from the source of interest (e.g., from a tumor), and is used to synthesize cRNA targets which are labeled (fluorescently, or by biotin which is later used for attachment of a fluorescent tag). The labeled cRNA is then hybridized to the microarray.

The surface of the microarray is scanned with a laser scanning device that measures the fluorescence intensity at each position in the

microarray. The fluorescence intensity of each spot on the array is proportional to the amount of cRNA which is hybridized to that spot.

Calculation of Average Difference

We use the *average difference* as our measure for the expression level of each gene (Lipshutz et al., 1999). The average difference is average of the differences of the intensities between each PM_i and its MM_i:

$$D_i = PM_i - MM_i$$

Twenty such D-values are calculated for each gene. For the calculation of the Average Difference we discard the most extreme *D* values and calculate the Average Difference as follows:

$$AvgDiff = \frac{1}{N}\sum_{i=1}^{N} D_i$$

N is the number of probe pairs included in the calculation.

In chapter 3 we use *Frequency* values to measure genes expression levels. The *Frequency* values are directly proportional to the Average Difference values. Using spikes of bacterial genes at known concentrations, a standard curve is constructed for each chip, which is then used to interpolate all Average Difference values in order to calculate gene frequencies. The gene frequency is expressed as: gene copies per million transcripts. Since the frequency is calculated for each gene on every chip, data across many different experiments, at many different times can be compared. In contrast, using the Average Difference value does not allow a historical comparison of chip data (without scaling or normalization) because Average Difference values change from day to day, depending on many different factors.

Analysis Methods

Preprocessing and filtering

We start out with an expression matrix the data is organized in n_S columns (experiments) and n_g rows (genes on the chip). Denote by A_{gs} the "average difference" of gene *g* in sample *s*. First, we threshold the data; we set $B_{gs} = A_{gs}$ for sizeable values, $A_{gs} \geq T$, and replaced low values, $A_{gs} < T$, by $B_{gs} = T$, *T* being our chosen threshold. Next, log is taken: $E_{gs} = \log_2 B_{gs}$, and the genes are filtered on the basis of their variation across the samples. Denote by $\overline{E}_g$ the average of the E_{gs} values obtained for gene *g* over all n_s samples and by $\sigma_g = \sqrt{\sum_s (E_{gs} - \overline{E}_g)^2 \Big/ (n_s - 1)}$ their standard deviation. Only those genes that satisfied $\sigma_g > \sigma_T$ are used for the analysis. σ_T is ourhosen threshold for the standard deviation.

Normalization Prior to Clustering

Before each clustering iteration the rows of the data matrix (genes) are centered (mean=0) and normalized to standard deviation of 1:

$$E'_{gs} = \frac{E_{gs} - \overline{E}_g}{\sqrt{\sum_s (E_{gs} - \overline{E}_g)^2}}$$

Supervised Analysis

We used supervised analysis (hypothesis testing) to identify genes, one at a time, whose expression differ between the two known classes *A, B* of n_A and n_B samples, respectively.

The Wilcoxon (Mann Whitney) Rank Sum test

We used the Wilcoxon Rank Sum test (Mann Whitney test) (see Lowry, 2003) in order to find genes differentially expressed between the two

groups of samples. We chose to use the Rank Sum test because it is a non-parametric test, so it does not assume the underlying source population(s) to be normally distributed.

The only assumptions of the Rank-Sum test are: (1) That the two samples are randomly and independently drawn; (2) That the dependent variable is intrinsically continuous (3) That the measures within the two samples have the properties of an ordinal scale of measurement (e.g. there is an order relation on the elements).

The Rank-Sum test is very simple and involves the assignment of ranks to the pulled values of both samples. We use the sum of ranks of the small group as our statistic, *W*, in order to obtain a p-value. This p-value is the probability to get such a rank distribution, or a better one, between the two sample groups, under the null hypothesis that the values of both samples come from the same distribution.

Given two groups of samples A and B, with n_A and n_B sample sizes respectively, $n_A + n_B = N$, and $n_A < n_B$.

The procedure is as follows:

For Small sample size (n_A<10, or N<20):

1) Combine the two sample groups.
2) Assign ranks from the lowest value to the highest 1,2…N. If there are ties the tied values are assigned the average of the ranks.
3) The statistic W_A is the sum of the ranks of the smaller sample.
4) All the possibilities for partitioning the $n_A + n_B$ samples into group A' with n_A samples and group B' with the remaining n_B samples are considered and for each such partitioning *W* is calculated. Denote by *m* the number of W values that are more extreme than W_A (the number of values which are larger or smaller than W_A, whichever turnsout to be less). Denote the number of permutations N_P, the number of such permutations is $N_p = \frac{N!}{n_A! n_B!}$. Calculate the p-value: $p = \frac{2 \times m}{N_P}$ (multiply by 2 because this is a two sided test).

For Large sample size (n_A>10, N>20):: W_A is calculated in the same manner as before. If the sample size is large enough (n_A>10, N>20) then W is approximately normally distributed with a population mean of $\mu = \frac{n_A(N+1)}{2}$ and standard deviation of $\sigma = \sqrt{\frac{n_B \times n_A(N+1)}{12}}$. Therefore it is possible to calculate a Z score: $Z = \frac{W_A - \mu}{\sigma}$ and obtain a p-value, using the Z distribution.

Two Way ANOVA

The two-way analysis of variance is a procedure that examines the effects of two independent variables in parallel; in addition, two-way ANOVA allows you to determine whether the two independent variables **interact** with respect to their effect on the dependent variable (Lowry, 2003).

Typically, the sampled values are organized in a two-way table, the rows corresponding to the different levels of one variable and the columns correspond to the levels of the other. Thus, there are $n_R \times n_C$ groups of samples, n_R being the number of rows and n_C the number of columns.

Denote the two independent variables "A" (rows) and "B" (columns), the two-way ANOVA tests 3 null hypotheses simultaneously: (i) factor "A" does not affect the dependent variable, so the values of both rows are taken from the same distribution; (ii) factor "B" does not affect the dependent variable, so the values of both columns are taken from the same distribution; (iii) the two independent variables do not interact with respect to their effect on the dependent variable, so the mean of the values of each group, m_{g*}, is a simple additive combination of the row mean m_R and the column mean m_C.

Similarly to the one-way ANOVA (Lowry, 2003), the two-way ANOVA uses the sum squared deviates (SS, Sum of Squares) from the

calculated *expected* means, to measure the aggregate degree of difference among the groups. The measure of random variability that exists inside each group is measured by the sum of squared deviates of the values inside each group from the group mean.

Thus, in order to measure the effect of factor "A" we calculate the true row means, and the total mean, m_T, of the values. Under the null hypothesis that the values in both rows are taken from the same distribution, the expected mean is m_T. Therefore SS_{Rows} is calculated the following way:

$$SS_{Rows} = N_{R1}(m_{R1} - m_T)^2 + N_{R2}(m_{R2} - m_T)^2 + ... + N_{R2}(m_{R2} - m_T)^2$$

Where N_{Ri} is the number of samples in row *i*, m_{Ri} is the mean of the values in row *i*, and m_T is the total mean of the entire set of values.
The SS_{Rows} value is now divided by the degrees of freedom to give a value denoted MS_{Rows}. The degrees of freedom are given by: $df = n_R - 1$.

In a similar way, we measure the effect of factor "B" which is measured by SS_{Cols}:

$$SS_{Cols} = N_{C1}(m_{C1} - m_T)^2 + N_{C2}(m_{C2} - m_T)^2 + ... + N_{Ci}(m_{Ci} - m_T)^2$$

Where N_{Ci} is the number of samples in column *i*, and m_{Ci} is the mean of the values in column *i*. SS_{Cols} value is divided by the degrees of freedom to give a value denoted MS_{Cols}. The degrees of freedom are given by: $df = n_C - 1$.

The third null hypothesis is that the two independent variables do not interact with respect to their effect on the dependent variable. In the case of no interaction the mean of any particular one of the individual groups, m_{gi} should be a simple additive combination of m_{r*} and m_{c*}, *

is the group's corresponding row or column number. Therefore, the expected mean according to the null hypothesis is:

$$em_{gi} = m_{r*} + m_{c*} - m_T$$

Therefore the SS_{int} is calculated as follows:

$$SS_{\text{int}} = N_{g1}(m_{g1} - em_{g2})^2 + N_{g2}(m_{g2} - em_{g2})^2 + ... + N_{gi}(m_{gi} - em_{gi})^2$$

Where N_{gi} is the number of samples in group *i*, m_{gi} is the observed mean of the values in group *i*, and em_{gi} is the expected mean of group *i*. SS_{int} value is divided by the degrees of freedom to give a value denoted MS_{int}. The degrees of freedom are given by: $df = (n_R - 1)(n_C - 1)$.

In addition to these three values we calculate the SS within groups, SS_{wg}, as our estimate of the random variability which exists inside the groups.

$$SS_{wg} = \sum_i \sum_j (x_{ij} - m_{gi})^2$$

Where *i* is the group number (1… $n_R \times n_C$) and *j* is the element number in group *i* (1… N_{gi}). x_{ij} is the *j*-th element in group *i*. SS_{wg} value is divided by the degrees of freedom to give a value denoted MS_{wg}. The degrees of freedom are given by: $df = N_T - n_R n_C$. Where N_T is the total number of elements in our entire set.

The statistic used in the two-way ANOVA is the *F ratio*, the same statistic used in the one-way ANOVA, in this case we calculate three *F ratios:*

$$F_{Rows} = \frac{MS_{Rows}}{MS_{WG}}$$

$$F_{Cols} = \frac{MS_{Cols}}{MS_{WG}}$$

$$F_{\text{int}} = \frac{MS_{\text{int}}}{MS_{WG}}$$

For each *F-ratio* we obtain a p-value from standard tables. Thus for each of the three null hypotheses we have an individual p-value and we can reject each null hypothesis individually.

Controlling the False Discovery Rate (FDR)

The Multiplicity Problem

Usually, when performing a statistical test, the null hypothesis is rejected at a level of $p<0.05$ ($p < \alpha, \alpha = 0.05$). Traditionally, the "$p < 0.05$" rule was formulated for a single test with the following logic: if $p < 0.05$ was observed, than there is a *1 in 20* chance that the null hypothesis holds. A *1 in a 20* chance is quite a small chance, and thus, one can reject the null hypothesis (of no real effect), and accept the conclusion that the effect is real.

This logic does not hold when more than one test is preformed in a single study. When multiple tests are preformed, an event of p<0.05 is expected to occur in 1 test from about every 20 tests carried out, even if the null hypothesis is true for all cases. In a typical DNA chip experiment, thousands of genes are being tested, thus for many of them the null hypothesis is falsely rejected (we will refer to them as "false positives"). In our case, we perform such tests on *N*=3000 genes. At the level of $p<0.05$ one expects (for random data) to "find" around 150 genes that will be falsely identified as separating the two groups of samples.

The False Discovery Rate (FDR)

It is clear that the smaller the p-value for which we reject the null hypothesis, the smaller the number of false positives we obtain. Therefore one way to deal with the false positives is to reject the null hypothesis for lower p-values. On the other hand, the more we reduce the α, for which we reject the null hypothesis, the more true cases (cases for which the null hypothesis is wrong) we miss.

The multiplicity problem was originally addressed by methods to control the family-wise type I error rate (FEW) which is the probability of committing at least one error in the family of hypotheses. A simple example of FEW is the Bonferroni method. Using this method, we reject the null hypothesis only in cases where $p < \frac{\alpha}{N}$, N being the number of tests preformed. This insures that the expectancy of false positives is α, and thus the probability to get even one false positive is less than α.

In DNA microarray experiments, the number of tests preformed is in the order of thousands. Therefore, a method such as Bonferroni will require very small p-values and will result in a significant loss of power. As an alternative, one can supply a measure for the expected proportion of falsely discovered genes among the list of genes that are identified; the expected proportion is the FDR (Benjamini and Hochberg, 1995).

Let R denote the number of hypotheses rejected by the procedure, V the number of true null hypotheses that are wrongly rejected. Then:

$$FDR = E\left(\frac{V}{R}\right)$$

FDR places a weaker demand on the genes identified (the probability for each gene to be falsely rejected is controlled, as opposed to the probability of no false positives at all, as in Bonferroni). Therefore less true genes are rejected.

The Procedure

Let N be the number of null hypotheses tested. For each hypothesis H_g, a test statistic is calculated with a corresponding p-value, p_g.

The N genes are ordered according to their p_g values. An upper bound, q, for the fraction of false positives is set; and the minimal index, j, for which $p_i > i \times q/N$ is found for all $i > j$. The null hypothesis is rejected for all genes with index $i \leq j$. At the end of this procedure we are left with a list of genes for which the expected fraction of false positives is q. (See Benjamini and Hochberg, 1995 for further information regarding FDR).

Unsupervised Analysis

Clustering Analysis

Clustering is the process of partitioning elements into groups such that "similar" elements will be grouped together and "different" elements will not be in the same group. Note that this definition is fuzzy because there are many definitions for the clustering problem. The algorithm we'll be using is the Super-paramagnetic clustering algorithm.

Super Paramagnetic Clustering (SPC)

SPC is based on the physical properties of an inhomogeneous ferromagnet (Blatt et al., 1996). SPC uses a particular cost function for each partition and generates an ensemble of partitions at a fixed value of the average cost (average over the ensemble). The SPC cost function uses a distance function between the elements, and penalizes assignment of close elements to different partitions. The probability for a given partition configuration is given by the Gibbs distribution where the temperature defines the average cost. At every temperature the probability that a pair of elements is assigned to the same partition is calculated, by averaging over all the different partition configurations at that temperature, according to their probabilities. Elements will be assigned to the same cluster only if they appear with a high enough probability in the same partition. Hence, for each temperature we have a different natural configuration of clusters. A stable cluster is a cluster that "lives" and does not separate into different groups for a large range ΔT.

The advantages of SPC are stability against noise, generating a hierarchy seen as a dendrogram ("tree view") and providing a way to recognize stable clusters, using a single distance function between the elements. In addition SPC does not need specification of the number of clusters in advance, a major advantage once working with large

data sets, as microarray data. In particular, SPC provides a reliable stability index for clusters.

For both genes and samples the algorithm uses the Euclidean distance to measure the distance between the normalized elements (see preprocessing for pre-clustering normalization).

Coupled Two-Way Clustering (CTWC)

Coupled Two-Way Clustering (Getz et al., 2000) is a method for reducing noise by focusing on small subsets of genes and samples. This is achieved by using only genes (and samples) that were identified previously as a stable cluster for the clustering process. The procedure is iterative; each stable cluster of genes that was found is used, in the next iteration, for the clustering of each of the stable clusters of samples that were previously found and vice versa. This process is repeated until no more clusters answering our criteria are discovered.

This method is especially appropriate for analyzing DNA microarray data, where different biological processes are simultaneously occurring, influencing different genes and samples. As a result, a great deal of noise is present, masking the effects individual processes. By focusing on correlated groups of genes and samples CTWC is able to minimize the noise generated by the majority of genes and identify specific biological processes involving specific genes or samples. Moreover, by using a group of correlated genes, noise of the individual measurements averages out and is reduced.

CTWC can be used with a variety of clustering algorithms. We use CTWC with SPC because of its robustness against noise and because it is one of the few algorithms able to provide a stability index to each cluster.

Reference List

Benjamini,Y. and Hochberg,Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. *57*, 289-300.

Blatt,M., Wiseman,S., and Domany,E. (1996). Superparamagnetic Clustering of Data. Physical Review Letters *76*, 3251-3254.

Getz,G., Levine,E., and Domany,E. (2000). Coupled two-way clustering analysis of gene microarray data. Proc. Natl. Acad. Sci. U. S. A *97*, 12079-12084.

Lipshutz,R.J., Fodor,S.P., Gingeras,T.R., and Lockhart,D.J. (1999). High density synthetic oligonucleotide arrays. Nat. Genet *21*, 20-24.

Lockhart,D.J., Dong,H., Byrne,M.C., Follettie,M.T., Gallo,M.V., Chee,M.S., Mittmann,M., Wang,C., Kobayashi,M., Horton,H., and Brown,E.L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. *14*, 1675-1680.

Lockhart,D.J. and Winzeler,E.A. (2000). Genomics, gene expression and DNA arrays. Nature *405*, 827-836.

Lowry, R. Concepts and Applications of Inferential Statistics. http://faculty.vassar.edu/lowry/webtext.html . 2003.
Ref Type: Electronic Citation

Chapter 6

Summary

DNA microarray data was analyzed using advanced clustering and statistical methods. This work involved two major projects, bone development and leukemia. In addition a small dataset of mutant p53 experiments was analyzed. The methods included statistical tests such as Ranksum, and two-way ANOVA, FDR control for multiple tests, SPC clustering and CTWC clustering method.

Leukemia DNA chip data were obtained from patients of AML and ALL, with or without MLL chromosome translocations, In addition there were 7 cell-line samples. The MLL gene is a gene commonly involved in acute leukemias. Using various statistical methods we were able to identify specific genes associated with the MLL translocation in ALL and in AML. In addition, we were able to identify genes more specifically associated with differentiation stages in leukemia, or with CD10- tumors. After the isolation of genes associated with these specific attributes, we were able to propose an interpretation to the fact that ALL patients with MLL translocations which differ from the rest of the ALL patients with MLL chromosome translocation; namely, that the difference is due to genes sensitive to differentiation (and CD10- genes), but the 3 have the same gene expression profile of the genes involved specifically in MLL translocation. This finding may suggest a new distinct subfamily of ALL with MLL translocation, which possibly have a different differentiation state. In addition, by using two-way ANOVA we were able to find genes which display behaviors that may be connected to particular regulation mechanisms. CTWC of the data revealed one unsuspected partition into primary tumors and cell-lines, and another partition of the ALLs separating into a stable cluster of ALLs bearing the MLL translocation. The genes that were found to separate the ALLs in this way, were not found in the statistical analysis. These

results emphasize the need to use both supervised and unsupervised methods in the analysis of DNA microarray data.

In the second project we analysed DNA microarray data of the developing bone. The dataset contained timecourse (7 timepoints: day 1, day 3, day 5, day 7, day 10, day 15, and day 20) measurements from a control group and an experimental group developing bone. We used two-way ANOVA in order to screen the genes that have a difference of behavior between the control group and the experimental group over time. These genes were then clustered using SPC in order to assign them to groups with similar profiles. We focused on several clusters, two of which peak in the samples obtained at the middle of the experiment (at the midst of the bone formation process). These two clusters contain, in addition to many known genes relevant to the process, genes involved in the Wnt pathway. The Wnt signaling pathway plays a crucial role in embryonic development and cell differentiation but does not have any known role in skelatogenesis. In addition to these genes, activins and their inhibitors were also upregulated, suggesting an involvment for activin in the bone formation process. In addition, two clusters of genes were found to be upregulated in the late stages of the experiments. These genes contained immunoglubolin chains, maybe corresponding to the formation of bone marrow and the subsequent invasion of B cells, and degradation enzymes for extracellular matrix modulation. Given the amount of meaningful results we received and their consistency throughout several datasets, we conclude that these combined methods constitute powerful profile analysis tools for timecourse chip experiments. In addition CTWC was used. One gene cluster corresponding to many of the genes that were found also in the supervised analysis was found to separate the experiment samples from the control group, and to order them in chronological order.

Mutations in the tumor suppressor p53 are found in more than 50% of the cancers. The mechanism by which mutation in p53 leads to cancer is believed to be not only due to the loss of function of p53, but

also due to a gain of function mechanism. We used permutation analysis to obtain genes whose expression levels were different between cells lacking p53 and cells carrying one of two different p53 mutants. Due to the small number of chips we could not identify specific genes, but at the lowest p-value we found 200 genes. The number of such genes expected from a random dataset of the same size is 100, thus indicating an activity of the mutant p53. In addition to the supervised permutation test, we preformed an unsupervised clustering analysis. The genes were filtered in a way that only genes acting consistently within identical repeats of experiments were taken. Clustering of these genes showed that almost all of them belong to two distinct groups, exhibiting different behaviors between the cells lacking p53 and cells containing the mutant p53, again indicating an activity for both mutants.

Supervised and unsupervised methods complement each other in the analysis of DNA microarray data. While supervised methods use prior knowledge about the data and provide tools for the search of genes separating known groups of samples, the unsupervised methods discover unexpected partitions in the data. Clustering also identifies genes which show a correlated behavior, whereas statistical tests identify genes which have different distribution of expression levels between the sample groups tested. Finally, utilizing these two methods together can help extract the information underlying in the data.

Appendix A

Table 1 | Genes Differentially Expressed between ALL with t(4:11) and ALL Without.. FDR =0.05

OVEREXPRESSED in t(4:11)

Num	Previously Found	Gene ID	Accession	Symbol and function	p-value
1		36873_at	D16532	VLDLR, very low density lipoprotein receptor; signal transduction, modulation of Dab1/Tau phosphorylation	0
2	RY	40763_at	U85707	MEIS1, myeloid ecotropic viral integration site 1 homolog (mouse); homeobox protein, murine leukemia	0
3	R	41448_at	AC004080	HOXA10; homeo box A10, sequence specific transcription factor	0.000022
4	YA	33412_at	AI535946	LGALS1, lectin, galactoside-binding, soluble 1 (galectin1); cell apoptosis and differentiation	0.000024
5		37479_at	M54992	CD72 antigen; B cell proliferation and differentiation	0.000037
6	RYA	37809_at	U41813	HOXA9; homeo box A9, sequence specific transcription factor, murine myeloid leukemia	0.000041
7	A	31472_s_at	AF098641	CD44, CD44 isoform (Indian blood group system)	0.000056
8	YA	1126_s_at	L05424	CD44, CD44 antigen (Indian blood group system); cell surface receptor, lymphocyte activation/homing	0.000056
9		35236_g_at	AA099265	RECK, reversion-inducing-cystein-rich protein with kazal motifs; membrane glycoprotein, tumor suppression	0.000063
10	A	31575_f_at	M14087	HL14, beta-galactoside-binding lectin	0.000068
11		40784_at	Z69030	PPP2R5C, protein phosphatase 2, regulatory subunit B (B56), gamma isoform	0.000069
12	YA	2036_s_at	M59040	CD44, CD44 antigen (Indian blood group system); cell surface receptor, lymphocyte activation/homing	0.000069
13		31880_at	D83767	D8S2298E (reproduction 8); fertilization	0.000086
14		37251_s_at	AF016004	GPM6B, glycoprotein M6B; integral membrane protein, expressed in neurons and glia	0.000095
15	A	38004_at	X96753	CSPG4, chondroitin sulfate proteoglycan 4 (melanoma-associated)	0.000097
16	A	37978_at	D78177	QPRT, quinolinate phosphoribosyltransferase; biosynthesis of NAD and NADP	0.000104
17		37724_at	V00568	MYC, v-myc myelocytomatosis viral oncogene homolog (avian); transcription factor, cell proliferation	0.000126
18	A	32184_at	X61118	LMO2, LIM domain only 2 (rhombotin-like 1); transcription factor, red blood cell development	0.000126
19		40785_g_at	Z69030	PPP2R5C, protein phosphatase 2, regulatory subunit B (B56), gamma isoform	0.000153
20		39210_at	M58597	FUT4, fucosyltransferase 4 (alpha (1,3) fucosyltransferase, myeloid-specific); glycosylation	0.000187
21	A	36753_at	AF072099	immunoglobulin-like transcript 3 protein, variant 1	0.000207
22		39244_at	M28211	RAB4A, member RAS oncogene family; protein transport	0.000224
23		37193_at	D78335	UMPK, uridine monophosphate kinase	0.000224
24		1452_at	U24576	LMO4, LIM domain only 4; transcription factor	0.000226
25		36053_at	AF041248	CDKN2C, cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4); negative control of cell proliferation	0.000226
26		37913_at	J00140	DHFR, dihydrofolate reductase; glycine and purine synthesis	0.00025
27		37558_at	U97188	KOC1, IGF-II mRNA-binding protein 3	0.000267
28		37184_at	L37792	STX1A, syntaxin 1A (brain); neurotransmission	0.000273
29		33358_at	W29087	KIAA1157	0.000273
30		1753_s_at	AD000092	RAD23A, RAD23 homolog A (S. cerevisiae); DNA excision repair	0.000336
31		36184_at	L06419	PLOD, procollagen-lysine, 2-oxoglutarate 5-dioxygenase (lysine hydroxylase, Ehlers-Danlos syndrome)	0.00038
32		176_at	U37352	PPP2R5C, protein phosphatase 2, regulatory subunit B (B56), gamma isoform	0.000396
33	A	39732_at	X73882	MAP7, microtubule-associated protein 7	0.000416
34	RA	41470_at	AF027208	PROML1, prominin-like 1 (mouse), AC133, CD133 antigen; coexpressed with CD34 on blood progenitors	0.000457
35		36313_at	M55267	EVI2A, ecotropic viral integration site 2A; membrane protein	0.000467
36		36229_at	U58917	IL17R, interleukin 17 receptor	0.000472

37		39091_at	AF070523	JWA; vitamin A responsive, cytoskeleton related	0.000472
38	A	1327_s_at	U67156	MAP3K5, mitogen-activated protein kinase kinase kinase 5; induction of apoptosis, activation of JUN	0.00051
39	A	36312_at	L40377	SERPINB8, serine (or cystein) proteinese inhibitor, clade B (ovalbumin), member 8	0.000548
40		38296_at	AL050196	DKFZP586D2223	0.000566
41	A	41191_at	AB023209	KIAA0992, palladin	0.000569
42	A	37743_at	U60060	FEZ1, fasciculation and elongation protein zeta1 (zygin I)	0.000594
43	A	33936_at	D86181	GALC, galactosylceramidase (Krabbe disease); lysosomal catabolism of major lipids	0.000618
44		34262_at	Y15909	DIAPH2, diaphanous homolog 2 (Drosphila); oogenesis	0.000634
45		40701_at	U75362	ISP13, ubiquitin specific protease 13 (isopeptidase T-3)	0.00068
46		31900_at	U33429	KCNAB2, potassium voltage-gated channel, shaker-related subfamily, beta member 2; K+ channel regulator	0.000681
47		1973_s_at	V00568	MYC, v-myc myelocytomatosis viral oncogene homolog (avian); transcription factor, cell proliferation	0.000682
48	RYA	37810_at	U82759	HOXA9; homeo box A9, sequence specific transcription factor, murine myeloid leukemia	0.000732
49	A	38223_at	AB024057	VRP, vascular Rab-GAP/TBC-containing	0.000732
50		41270_at	AA019936	SC65, nucleolar autoantigen (55kD) similar to rat synaptonemal complex protein	0.000776
51	Y	32215_i_at	AB020685	KIAA0878	0.000783
52		40019_at	M60830	EVI2B, ecotropic viral integration site 2B; membrane protein	0.000812
53		35136_at	AL031387	P15-2, hypothetical protein P15-2	0.000812
54		36922_at	X59618	RRM2, ribonucleotide reductase M2 polypeptide; synthesis of DNA precursors	0.000933
55		131_at	X83928	TAF21, TATA box binding protein (TBP)-associated factor, RNA polymerase II, I, 28 kD; TFIID complex	0.00095
56		1178_at	J00140	DHFR, dihydrofolate reductase, alternative splice 6	0.000961
57		36599_at	M55905	ME2, malic enzyme 2, NAD(+)-dependent, mitochondrial; pyruvate metabolism	0.000964
58		33847_s_at	AI304854	CDKN1B, cyclin-dependent kinase inhibitor 1B (p27, Kip1); cell cycle control, cell cycle arrest	0.000968
59		392_g_at	X89416	PPP5C, protein phosphatase 5, catalytic subunit	0.001036
60	YA	757_at	D28364	ANXA2, annexin A2	0.00104
61		36491_at	D82345	TMSNB, neuroblastoma thymosin beta; binds actin monomers, cytoskeleton organization	0.001052
62		41401_at	U57646	CSRP2, cysteine and glycine-rich protein 2; development of the embryonic vascular system	0.001141
63		37009_at	AL035079	CAT, catalase; protection from hydrogen peroxide	0.001149
64		40587_s_at	AF054186	EEF1E1, eukaryotic translation elongation factor 1 epsilon 1	0.001149
65	A	1914_at	U66838	CCNA1, cyclin A1; cell cycle control	0.00123
66		39018_at	AF026977	MGST3, microsomal glutathione S-transferase 3; membrane protein	0.001236
67		33382_at	M92449	ASAHL, N-acylsphingosine amidohydrolase (acid ceramidase)-like	0.001282
68		33405_at	N90755	CAP2, adenylyl cyclase-associated protein 2	0.0013
69		33126_at	L13435	chromosome 3p21.1 gene	0.001357
70		508_at	U43923	SUPT4H1, suppressor of Ty 4 homolog 1 (S. cerevisiae); transcription initiation, chromatin modelling	0.001359
71	A	873_at	M26679	HOXA5, homeo box A5; sequence specific transcription factor	0.001374
72	A	31684_at	M62896	ANXA2P1, annexin A2 pseudogene 1 (lipocortin 2)	0.001418
73	YA	36777_at	AJ001687	NKG2D, natural killer gene complex receptor; immunological response	0.001583
74		37747_at	U05770	ANX5, annexin A5; inhibitor of blood coagulation	0.001599
75	A	35985_at	AB023137	AKAP2, A kinase (PRKA) anchor protein 2	0.001607
76		37471_at	U94317	RPP40, ribonuclease P, 40 kD subunit	0.001609
77	YA	39338_at	AI201310	S100A10, S100 calcium binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11)	0.001609
78		39685_at	AL050282	E46L, like mouse brain protein E46	0.001609
79		32791_at	L19183	MAC30 hypothetical protein	0.001609

80		40966_at	AF099989	SPAK, Ste-20 related kinase	0.001609
81		1771_s_at	J03278	PDGFRB, platelet-derived growth factor receptor, beta polypeptide; signal transduction	0.001746
82		394_at	X92106	BLMH, bleomycin hydrolase; proteolysis and peptidolysis	0.001746
83		36149_at	D78014	DRYSL3, dihydropyrimidinase-like 3	0.001778
84		32615_at	J05032	DARS, aspartyl-tRNA synthetase	0.001886
85		35974_at	U10485	LRMP, lymphoid-restricted membrane protein	0.001895
86		855_at	S78085	PDCD2, programmed cell death 2	0.001895
87		40786_at	U37352	PPP2R5C, protein phosphatase PP2A, regulatory subunit B (B56), gamma isoform	0.001898
88		35255_at	AF098799	RANBP7, Ran binding protein 7	0.001898
89	YA	2062_at	L19182	IGFBP7, insulin-like growth factor binding protein 7	0.001898
90		205_g_at	M74297	HOXA4, homeo box A4; sequence specific transcription factor	0.002014
91		31655_at	AL031737	DNA sequence from clone 8B22 on chromosome 1p35.1-36.21, similar to cytoplasmic dynein light chain 1	0.002057
92		37483_at	AB018287	HDAC9-PENDING, histone deacetylase 9	0.00222
93		37357_at	D00723	GCSH, glycine cleavage system protein H (aminomethyl carrier)	0.002231
94		39175_at	D25328	PFKP, phosphofructokinase, platelet; glycolysis	0.002231
95	A	38201_at	U21551	BCAT1, branched chain aminotransferase 1, cytosolic; catabolism of leucine, isoleucine and valine	0.002235
96		33267_at	AF035315	clone 23664 and 23905	0.002235
97		33891_at	AL080061	CLIC4, chloride intracellular channel 4	0.002235
98		36631_at	D49396	PRDX3, peroxiredoxin 3; protection from oxidative damage	0.002235
99		40634_at	M86667	NAP1L1, nucleosome assembly protein 1-like 1	0.002235
100		41356_at	W27619	BCL11A, B-cell CLL/lymphoma 11A (zinc finger protein)	0.002235
101		1854_at	X13293	MYBL2, v-myb myeloblastosis viral oncogene homolog (avian)-like 2; transcription factor	0.002235
102		40676_at	U37139	ITGB3BP, integrin beta 3 binding protein (beta 3-endonexin); signal transduction	0.002408
103	A	32475_at	AF025529	LILRA1, leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 1	0.002585
104		36203_at	X16277	ODC1, ornithine decarboxylase 1, polyamine biosynthesis	0.00262
105		40081_at	L26232	PLTP, phospholipid transfer protein	0.002624
106		40467_at	AB006202	SDHD, succinate dehydrogenase complex, subunit D; membrane protein, tricarboxylic acid cycle	0.002624
107		34376_at	AB019517	PKIG, protein kinase (cAMP-dependent, catalytic) inhibitor gamma	0.002624
108		777_at	D13988	GDI2, GDP dissociation inhibitor 2; regulation of GDP/GTP exchange of RAB proteins	0.002624
109		455_at	U66618	SMARCD2, SWI/SNF related, matrix associated, actin dependent regulator of chromatin	0.002624
110		36688_at	U11313	SCP-2, sterol carrier protein 2; lipids transfer between membranes	0.002829
111		1803_at	X05360	CDC2, cell division cycle 2, G1 to S and G2 to M; cell cycle control, mitotic start control point	0.00292
112		34106_at	L01694	GNA12, guanine nucleotide binding protein (G protein) alpha 12; signal transduction	0.003034
113		2033_s_at	U10564	WEE1, wee1 + homolog (S. pombe); protein tyrosine kinase, cell cycle control	0.003057
114		33800_at	AF036927	ADCY9, adenylate cyclase type 9; signal transduction	0.003071
115		1599_at	L25876	CDKN3, cyclin dependent kinase inhibitor 3 (CDK2-associated dual specificity phosphatase); cell cycle arrest	0.003071
116	Y	854_at	S76617	BLK, B lymphoid tyrosine kinase	0.003071
117		37985_at	L37747	LMNB1, lamin B1; component of the nuclear lamina	0.003075
118		33369_at	AI535653	SC4MOL, sterol-C4-methyl oxidase-like; cholesterol biosynthesis	0.003075
119		38755_at	X84709	FADD, Fas (TNFRSF6)-associated via death domain; induction of apoptosis	0.003075
120		39099_at	X97064	SEC23A, Sec23 homolog A (S. cerevisiae); transport from endoplasmic reticulum to the Golgi apparatus	0.003075
121		430_at	X00737	PNP, purine nucleoside phosphorylase	0.003075

122		36858_at	D25218	RRS1, homolog of yeast ribosome biogenesis regulatory protein	0.003161
123		37188_at	X92720	PCK2, phosphoenolpyruvate carboxykinase 2	0.003303
124		35883_at	X66079	SPIB, Spi-B transcription factor (Spi-1/PU.1 related); ETS family, sequence spscific transcription factor	0.003545
125		33772_at	L25124	PTGER4, prostaglandin E receptor 4 (sybtype EP4); G-protein coupled receptor	0.003545
126		37282_at	AJ000186	MAD2L1, MAD2 mitotic arrest deficient-like1 (yeast); mitotic check point	0.003576
127	Y	794_at	X62055	PTPN6, protein tyrosine phosphatase, non-receptor type 6; hematopoiesis, signal transduction	0.003576
128		37663_at	X70649	DDX1, DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 1; DEAD box helicase family	0.003591
129		1839_at	M31469	RAN, TC4 member RAS oncogene family; nucleocytoplasmic transport	0.003596
130		38277_at	M29550	PPP3CB, protein phosphatase 3 (formerly 2B), catalytic subunit, beta isoform (calcineurin A beta)	0.003698
UNDEREXPRESSED in t(4:11)					
131	A	1992_at	U46922	FHIT, fragile histidine triad; nucleotide metabolism, tumor suppressor	0.00001
132		39424_at	U70321	TNFRSF14, tumor necrosis factor receptor superfamily, member 14; lymphocyte activation	0.000012
133	A	37343_at	U01062	ITPR3, inositol 1,4,5-triphosphate receptor type 3; signal transduction, small molecule transport	0.000013
134		1261_i_at	M16594	GSTA2, glutathione S-transferase A2	0.000017
135		1067_at	U03858	FLT3LG, fms-related tyrosine kinase 3 ligand; stimulates proliferation of early hematopoietic cells	0.000024
136		39650_s_at	AB007895	KIAA0435	0.000037
137		37413_at	J05257	DPEP1, dipeptidase 1 (renal); renal metabolism of glutathione	0.000046
138	A	41266_at	X53586	ITGA6, integrin alpha 6; laminin receptor, critical structure role in the hemidesmosome	0.000056
139	A	307_at	J03600	ALOX5, arachidonate 5- lipoxygenase; biosythesis of leukotrienes	0.000056
140	A	182_at	U01062	ITPR3, inositol 1,4,5-triphosphate receptor, type 3; signal transduction, small molecule transport	0.000062
141		1483_at	L34059	CDH4, cadherin 4, type 1, R-cadherin (retinal); cell adhesion	0.000069
142		36008_at	AF041434	PTP4A3, protein tyrosine phosphatase type IVA, member 3	0.000085
143		40049_at	X76104	DAPK1, death-associated protein kinase 1; mediating interferon-gamma-induced cell death	0.000093
144		33440_at	U19969	TCF8, transcription factor 8; repression of transcription (interleukin-2)	0.0001
145		38631_at	M92357	TNFAIP2, tumor necrosis factor, alpha-induced protein 2	0.000104
146		1971_g_at	U46922	FHIT, fragile histidine triad; nucleotide metabolism, tumor suppressor	0.000104
147		40016_g_at	AB002301	KIAA0303	0.00011
148		539_at	S59184	RYK, RYK receptor-like tyrosine kinase	0.000135
149	A	41715_at	Y11312	PIK3C2B, phosphoinositide-3-kinase, class 2, beta polypeptide	0.000187
150		39385_at	M22324	ANPEP, alanyl (membrane) aminopeptidase (aminopeptidase N, M, CD13, p150); receptor for coronavirus	0.000216
151		1325_at	U59423	MADH1, mothers against decapentaplegic homolog 1 (Drosophila); receptor-regulated transcription	0.000253
152		33910_at	AL049338	DKFZp564P116	0.000269
153		32156_at	AF044968	PVRL2, poliovirus receptor-related 2 (herpesvirus entry mediator B)	0.000273
154		40953_at	S80562	CNN3, calponin 3, acidic; thin filament-associated protein, smooth muscle contraction	0.000289
155		40511_at	X58072	GATA3, GATA-binding protein 3; transcriptional activator of T cell receptor alpha and delta genes	0.000294
156		1963_at	U01134	FLT1, fms-related tyrosine kinase 1, vascular endothelial growth factor receptor	0.000299
157		35955_at	S80864	CYCL, cytochrome c-like antigen; tumor antigen	0.000329
158		39824_at	AI391564	PTP4A3, protein tyrosine phosphatase type IVA, member 3	0.000329
159	A	40729_s_at	Y14768	LTB, lymphotoxin beta (TNF superfamily, member 3)	0.000396
160	A	35164_at	AF084481	WFS1, Wolfram syndrome 1 (wolframin)	0.000436

161	A	37280_at	U59912	MADH1, mothers against decapentaplegic homolog 1 (Drosophila); receptor-regulated transcription	0.000436
162		39255_at	X02750	PROC, protein C (inactivator of coagulation factors Va and VIIIa); regulating blood coagulation	0.000472
163		34905_at	AA977136	GRIK5, glutamate receptor, ionotropic, kainate 5; synaptic neurotransmission	0.000476
164		36482_s_at	Y15724	ATP2A3, ATPase, Ca++ transporting, ubiquitous; calcium transportation within the cell	0.000476
165		32554_s_at	Y12781	TBL1, transducin (beta)-like 1	0.000532
166		37014_at	M33882	MX1, myxovirus (influenza virus) resistance 1, interferon inducible protein p78 (mouse)	0.000569
167		33993_at	M22919	MYL6, myosin, light polypeptide 6, alkali, smooth muscle and non-muscle	0.00057
168		35939_s_at	L20433	POU4F1, POU domain class 4, transcription factor1; synaptogenesis, axonogenesis	0.00057
169		41036_at	AL022314	IL2RB, interleukin 2 receptor beta; signal transduction, endocytosis	0.00057
170		32107_at	AL050173	C21orf25, chromosome 21open reading frame 25	0.000633
171		1674_at	M15990	YES1, v-yes-1 Yamaguchi sarcoma viral oncogene homolog 1; protein tyrosine kinase	0.000647
172		40155_at	D31883	ABLIM, actin binding LIM protein; actin cytoskeleton	0.000676
173	A	36650_at	D13639	CCND2, cyclin D2; cell cycle control	0.000677
174		222_at	S79639	EXT1, exostosis (multiple) 1; glycosyltransferase	0.000677
175		34789_at	S69272	SERPINB, serine (or cystein) proteinase inhibitor, clade B (ovalbumin), member 6; inhibits thrombin	0.00068
176		35961_at	AL049390	DKFZp586O1318	0.000682
177		37645_at	Z22576	CD69, CD69 antigen (p60, early T-cell activation antigen); receptor, lymphocyte proliferation	0.000682
178		41868_at	J04131	GGT1, gamma-glutamyltransferase 1; synthesis and degradation of glutathione	0.000692
179		34897_at	W26524	PPP4R2, protein phosphatase 4, regulatory subunit 2	0.000732
180		32978_g_at	U49187	C6orf32, chromosome 6 open reading frame 32	0.000764
181	A	37680_at	U81607	AKAP12, A kinase (PRKA) anchor protein (gravin) 12; PKA/PKC compartmentation	0.000764
182	A	31886_at	X55740	NT5, 5' nucleotidase (CD73); hydrolyzes extracellular nucleotides into permeable nucleosides	0.00078
183		39837_s_at	AC004877	PAC clone RP4-751H13from 7q35-qter	0.000786
184		36566_at	AJ222967	CTNS, cystinosis, nephropathic; cystinosin family	0.000812
185		36155_at	D87465	KIAA0275	0.000813
186		1895_at	J04111	JUN, v-jun sarcoma virus 17 oncogene homolog (avian); transcription factor	0.000813
187		38761_s_at	AA487755	FKBP9, FK506 binding protein 9 (63 kD)	0.000874
188		39569_at	U72849	EVPL, envoplakin; component of the cornified envelope of keratinocytes	0.000913
189		34904_at	S40369	GRIK5, glutamate receptor, ionotropic, kainate 5; neurotransmission	0.000961
190		41504_s_at	AF055376	MAF, v-maf musculoaponeurotic fibrosarcoma oncogene homolog (avian)	0.000968
191	A	41710_at	AL079277	LOC54103	0.001014
192		39154_at	AI952982	GADD45G, growth arrest and DNA-damage-inducible, gamma; stress-response, apoptosis	0.001102
193		38767_at	AF041037	SPRY1, sprouty homolog 1, antagonist of FGF signaling (Drosophila); signal transduction	0.001149
194		35780_at	AF035292	KIAA0657	0.001184
195		32583_at	J04111	JUN, v-jun sarcoma virus 17 oncogene homolog (avian); transcription factor	0.00123
196	A	33809_at	AL049933	GNAI1, guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1; signaling	0.001311
197	A	37420_i_at	AL022723	clone RP3-377H14, HLA-F, major histocompatibility complex, class I, F precursor	0.001361
198		38968_at	AB005047	SH3BP5, SH3-domain binding protein 5 (BTK-associated); signal transduction	0.001361
199		1110_at	M21624	TRD, T cell receptor, delta	0.001361
200		879_at	M30818	MX2, myxovirus (influenza virus) resistance 2 (mouse); dynamin family	0.001361

201	A	41346_at	AJ007583	LARGE, like-glycosyltransferase	0.001478
202		39781_at	U20982	IGFBP4, insulin-like growth factor binding protein 4; signal transduction, proliferation	0.001557
203		34210_at	N90866	CDW52, CDW52 antigen (CAMPATH-1antigen)	0.001607
204		39327_at	D86983	D2S448, melanoma associated	0.001609
205		1586_at	M35878	IGFBP3, insulin-like growth factor binding protein 3; signal transduction, proliferation	0.001839
206		1237_at	S81914	IER3, immediate early response 3	0.001886
207		34176_at	AF091087	LOC57228	0.001895
208		40712_at	D26579	ADAM 8, a disintegrin and metalloproteinase domain 8	0.001898
209		38147_at	AL023657	SH2D1A, SH2 domain protein 1A; Duncan's disease (lymphoproliferative syndrome); T cell signaling	0.002035
210		33871_s_at	J02876	FOLR2, folate receptor 2 (fetal); folate transport	0.002035
211		37006_at	AI660656	ESTs, highly similar to IGJ, human immunoglobulin J chain	0.002191
212		32007_at	W29045	unknown cDNA	0.002235
213		36493_at	M33552	LSP1, lymphocyte-specific protein 1	0.002235
214		32786_at	X51345	JUNB, jun B proto-oncogene; sequence specific transcription factor	0.002235
215		36118_at	AJ000882	NCOA1, nuclear receptor coactivator 1	0.002235
216		1105_s_at	M12886	TRB, T cell receptor, beta	0.002235
217		37956_at	U37519	ALDH3B2, aldehyde dehydrogenase 3 family, member B2	0.002383
218		36394_at	AB012293	LY6H, lymphocyte antigen 6 complex, locus H	0.002392
219		1737_s_at	M62403	IGFBP4, insulin-like growth factor binding protein 4; signal transduction, proliferation	0.002585
220	A	1007_s_at	U48705	DDR1, discoidin domain receptor family, member 1; tyrosine kinase, insulin receptor subfamily	0.002609
221		32112_s_at	AI800499	AIM1, absent in melanoma 1	0.002624
222		1097_s_at	L31584	CCR7, chemokine (C-C motif) receptor 7; G-protein coupled receptor	0.002624
223		38124_at	X55110	MDK, midkine (neurite growth-promoting factor 2); pleiotrophin family	0.002711
224	A	41136_s_at	Y00264	APP, amyloid beta (A4) precursor protein (protease nexin-II, Alzheimer disease); neuronal receptor	0.003057
225		40224_s_at	AB014585	KIAA0685	0.003071
226		41656_at	AF043325	NMT2, N-myristoyltransferase 2; protein modification	0.003075
227		38364_at	AF068197	BCE-1	0.003075
228		33705_at	L20971	PDE4B, phosphodiesterase 4B cAMP-specific (phosphodiesterase E4 dunce homolog, Drosophila)	0.003322
229	A	38578_at	M63928	TNFRSF7, tumor necrosis factor receptor superfamily, member 7	0.003568
230		33263_at	X67098	HSRTSBETA	0.003568
231		40775_at	AL021786	ITM2A, integral membrane protein 2A	0.003591
232		32977_at	U49187	C6orf32, chromosome 6 open reading frame 32	0.003596
233		40451_at	AL080203	POLE, polymerase (DNA directed) epsilon; DNA repair and chromosomal DNA replication	0.003596
234		40562_at	M69013	GNA11, guanine nucleotide binding protein (G protein), alpha 11 (Gq class); signaling	0.003596
235	YA	1389_at	J03779	MME, membrane metallo-endopeptidase (neutral endopeptidase, enkephalinase, CALLA, CD10)	0.003596
236		1185_at	D49410	IL3RA12, interleukin 3 receptor, alpha subunit, exon 12	0.003596
237		556_s_at	M96233	GSTM4, glutathione S-transferase M4	0.003596

R- Found previously by Rozovskaia et al. (Rozovskaia et al., 2001)
Y- Found previously by Yeoh et al. (Yeoh et al., 2002)
A- Found previously by Armstrong et al. (Armstrong et al., 2002)

Table 2 | Genes Differentially Expressed between AML with ALL-1 rearrangement and AML Without. FDR =0.1

OVEREXPRESSED in AML with ALL-1 rearrangement

Num	Gene ID	Accession	Symbol and function	p-value
1	35318_at	AB007944	KIAA0475	0.000171
2	41431_at	AB023153	ICK, intestinal cell kinase	0.000213

3	33912_at	Y13834	ZMPSTE24, zinc metalloproteinase (STE24 homolog, yeast)	0.000222
4	41260_at	U59321	DDX17, DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 17 (72 kD); helicase, RNA-dependent ATPase	0.000318
5	33162_at	X02160	INSR, insulin receptor; signal transduction	0.000325
6	34936_at	AB012130	SLC4A7, solute carrier family 4, sodium bicarbonate cotransporter, member 7	0.000369
7	34737_at	AF058718	COG5, component of oligomeric golgi complex 5	0.00037
8	37303_at	AF057160	ADPRTL1, ADP-ribosyltransferase (NAD+; poly (ADP-ribose) polymerase) -like 1	0.00037
9	32044_at	D13635	KIAA0010, ubiquitin-protein isopeptide ligase (E3)	0.000475
10	706_at	B93370	NR3C1, nuclear receptor subfamily 3, group C, member 1; regulation of gene expression	0.000606
11	37973_at	AB018256	SNX13, sorting nexin 13	0.000636
12	41480_at	AF029669	RAD51C, RAD51 homolog C (S.cerevisae)	0.000758
13	38353_at	AF042378	GCP3, spindle pole body protein; microtubule cytoskeleton	0.000771
14	1944_f_at	AF001359	MLH1, DNA mismatch repair protein, alternatively spliced	0.000928
15	36048_at	AB015342	HRIHFB2436; endocrine regulator	0.000978
16	40437_at	AL049944	DKFZp564G2022	0.000978
17	36498_at	AI936759	AP1S1, adaptor-related protein complex 1, sigma 1 subunit; clathrin coat assembly protein	0.00105
18	39166_s_at	D83174	SERPINH2, serine/cysteine proteinase inhibitor, clade H (heat shock protein 47), member 2; collagen synthesis	0.00105
19	40982_at	AA926957	FLJ10534	0.00123
20	1348_s_at	S79219	PCCA, propionyl Coenzyme A carboxylase, alpha polypeptide; fatty acids metabolism	0.00123
21	176_at	U37352	PPP2R5C, protein phosphatase 2, regulatory subunit B (B56), gamma isoform	0.00123
22	36935_at	M23379	RASA1, RAS p21 protein activator (GTPase activating protein) 1	0.001234
23	549_at	S80343	RARS, arginyl-tRNA synthetase	0.001234
24	36968_s_at	AL050353	OIP2, Opa-interacting protein 2	0.001551
25	38035_at	AF072928	MTMR6, myotubularin related protein 6	0.001551
26	41557_at	D29641	KIAA0052	0.001551
27	32081_at	AB023166	CIT, citron (rho-interacting, serine/threonine kinase 21)	0.00172
28	31869_at	AB014540	SWAP70, SWAP complex protein, 70 kD	0.001941
29	37985_at	L37747	LMNB1, lamin B1; inner nuclear envelope	0.001941
30	40086_at	D87450	KIAA0261	0.001941
31	39097_at	X63753	SON, SON DNA binding protein; sequence specific transcription factor	0.001941
32	2000_at	U26455	ATM, ataxia telangiectasia mutated; kinase family, signal transduction, cell cycle, DNA repair	0.001941
33	40732_at	D83243	NPAT, nuclear protein, ataxia telangiectasia locus	0.002156
34	39761_at	D87682	KIAA0241	0.002211
35	41248_at	AB014589	CSTF2T, likely ortholog of mouse variant polyadenylation protein CSTF-64	0.002413
36	37663_at	X70649	DDX1, DEAD/H (Asp-Glu-Ala-Asp/His) box popypeptide 1; helicase family	0.00242
37	40459_at	S69189	ACOX1, acyl-Coenzyme A oxidase 1, palmitoyl; energy pathways, lipid metabolism	0.002443
UNDEREXPRESSED in AML with ALL-1 rearrangement				
38	37192_at	U28389	EPB49, erythrocyte membrane protein band 4.9 (dematin); villin/gelsolin family, actin-bundling protein	0.000171
39	1008_f_at	U50648	PRKR, protein kinase, interferon-inducible double stranded RNA dependent; inhibition of protein synthesis	0.000171
40	31506_s_at	L12691	NTR3, neutrophil peptide-3	0.000213

41	39908_at	AF069735	TAF6L, TAF6-like RNA polymerase II, p300/CBP-associated factor (PCAF)-associated factor, 65 kDa	0.000222
42	37149_s_at	U95626	LTR, lactotransferrin; iron transport	0.000241
43	293_at	X74864	HPX42, homeo box protein	0.000241
44	37405_at	U29091	SELENBP1, selenium binding protein 1	0.000252
45	32052_at	L48215	HBB, hemoglobin, beta; oxygen transport	0.000265
46	35473_at	Z74615	COL1A1, collagen, type1, alpha 1; skeletal development, epidermal differentiation	0.000286
47	31687_f_at	M25079	HBB, hemoglobin beta; oxygen transport	0.000356
48	37285_at	X60364	ALAS2, aminolevulinate, delta-, synthase 2 (sideroblastic/hypochromic anemia); heme biosynthesis	0.00037
49	35601_at	L00022	IGHE, immunoglobulin heavy constant epsilon	0.000493
50	33530_at	M33326	CEACAM8, carcinoembryonic antigen-related cell adhesion molecule 8	0.000503
51	31525_s_at	J00153	HBA1, hemoglobin, alpha 1; oxygen transport	0.000606
52	38715_at	W04490	GYPB, glycophorin B (includes Ss blood group); minor sialoglycoprotein in erythrocyte membranes	0.000606
53	32821_at	AI762213	LCN2, lipocalin 2 (oncogene 24p3)	0.000632
54	31793_at	AL036554	DEFA1, defensin, alpha 1, myeloid-related sequence	0.000771
55	35919_at	J05068	TCN1, transcobalamin I (vitamin B12 binding protein, R binder family)	0.000784
56	40062_s_at	X58851	MYL4, myosin, light polypeptide 4, alkali, atrial, embryonic	0.000811
57	1962_at	M14502	ARG1, arginase, liver; arginine degradation in the urea cycle	0.000811
58	31930_f_at	X63096	RHCE, Rhesus blood group, CcEe antigens	0.000978
59	34961_at	M88282	TACTILE, T cell activation, increased late expression	0.001179
60	36980_at	U03105	B4-2, prolin-rich protein with nuclear targeting signal	0.001541
61	38051_at	X76220	MAL, mal, T-cell differentiation protein	0.001546
62	1115_at	M25897	PF4, platelet factor 4	0.001546
63	33516_at	V00505	HBD, hemoglobin, delta	0.001551
64	32786_at	X51345	JUNB, jun B proto-oncogene; sequence specific transcription factor, BZIP family, JUN subfamily	0.001551
65	39209_r_at	M54995	PPBP, pro-platelet basic protein (includes platelet basic protein, beta-thromboglobulin, and others)	0.001747
66	677_s_at	J04430	ACP5, acid phosphatase 5, tartrate resistant	0.001747
67	33336_at	M27819	SLC4A1, solute carrier family 4, anion exchanger, member 1 (erythrocyte membrane protein band 3)	0.001935
68	37054_at	J04739	BPI, bactericidal /permeability-increasing protein	0.001941
69	32035_at	M16942	HLA-DRB4, major histocompatibility complex, class II, DR beta 4	0.001941
70	1992_at	U46922	FHIT, fragile histidine triad gene; nucleotide metabolism, tumor suppressor	0.001941
71	1971_g_at	U46922	FHIT, fragile histidine triad gene; nucleotide metabolism, tumor suppressor	0.001941
72	1223_at	X66362	PCTK3, PCTAIRE protein kinase 3	0.002413
73	37026_at	AF001461	COPEB, core promoter element binding protein; Krueppel family of zinc-finger proteins	0.00242
74	33121_g_at	AF045229	RGS10, regulator of G-protein signalling 10; signal transduction	0.00242
75	1894_f_at	L27065	NF2, neurofibromatosis 2 tumor suppressor	0.00242

Table 3 | Genes Differentially Expressed between ALL with t(4:11) and ALL Without AND AML with ALL-1 translocation and AML without. Intersections of FDR=0.15

OVEREXPRESSED in translocations

Gene ID	Accession	Symbol	Title and Function	p-value
176_at	U37352	PPP2R5C	protein phosphatase 2, regulatory subunit B (B56), gamma isoform	0.000002
37663_at	X70649	DDX1	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 1; DEAD box helicase family	0.000015
35255_at	AF098799	RANBP7	RAN binding protein 7	0.000015
37985_at	L37747	LMNB1	lamin B1; component of the nuclear lamina	0.000024

131_at	X83928	TAF21	TAF21, TATA box binding protein (TBP)-associated factor, 28 kD; TFIID complex	0.000028
35318_at	AB007944	KIAA0475	KIAA0475 gene product	0.000029
40701_at	U75362	USP13	ubiquitin specific protease 13 (isopeptidase T-3)	0.000031
981_at	X74794	MCM4	MCM4 minichromosome maintenance deficient 4 (S. cerevisiae); DNA replication	0.000049
40090_at	AI797997	WBSCR20	Williams Beuren syndrome chromosome region 20	0.000049
34936_at	AB012130	SLC4A7	solute carrier family 4, sodium bicarbonate cotransporter, member 7	0.000061
40786_at	U37352	PPP2R5C	protein phosphatase 2, regulatory subunit B (B56), gamma isoform	0.000062
38984_at	AB007896	KIAA0436	putative L-type neutral amino acid transporter	0.000080
335_r_at	L21990	SF3A2	splicing factor 3a, subunit 2, 66 kD (SAP62); binding of SNRNP complex to the BPS in mRNA	0.000116
33247_at	U86782	POH1	26S proteasome-associated pad1 homolog	0.000130
379_at	AB006679	APACD	ATP binding protein associated with cell differentiation	0.000135
40086_at	D87450	KIAA0261	KIAA0261 protein	0.000141
36968_s_at	AL050353	OIP2	Opa-interacting protein 2	0.000147
41480_at	AF029669	RAD51C	RAD51 homolog C (S. cerevisiae); DNA repair	0.000148
40459_at	S69189	ACOX1	acyl-Coenzyme A oxidase 1, palmitoyl; energy pathways, lipid metabolism	0.000153
2012_s_at	U34994	PRKDC	protein kinase, DNA-activated, catalytic polypeptide; DNA repair, recombination	0.000166
37304_at	U35451	CBX1	chromobox homolog 1 (HP1 beta homolog Drosophila); chromatin remodelling	0.000181
1798_at	U41060	LIV-1	LIV-1 protein, estrogen regulated	0.000186
38687_at	AL050051	DKFZP566D193	DKFZP566D193 protein	0.000194
41557_at	D29641	KIAA0052	KIAA0052 protein	0.000246
842_at	U48251	PRKCBP1	protein kinase C binding protein 1	0.000254
34768_at	AL080080	TXNDC	thioredoxin domain containing	0.000266
36099_at	M69040	SFRS1	splicing factor, arginine/serine-rich 1 (splicing factor 2, alternate splicing factor)	0.000276
34737_at	AF058718	COG5	component of oligomeric golgi complex 5	0.000279
39328_at	M11058	HMGCR	3-hydroxy-3-methylglutaryl-Coenzyme A reductase; sterol synthesis	0.000329
1448_at	D00762	PSMA3	proteasome (prosome, macropain) subunit, alpha type, 3; proteinase complex	0.000336
39788_at	X81889	PKP4	plakophilin 4; junctional plaques	0.000348
1850_at	U07418	MLH1	mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli); BRCA1-associated complex	0.000391
39432_at	AF038662	B4GALT4	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 4; glycosylation	0.000421
37359_at	D14658	KIAA0102	KIAA0102 gene product	0.000455
38349_at	AF038564	ITCH	itchy homolog E3 ubiquitin protein ligase (mouse)	0.000611
32777_at	Y12478	WRB	tryptophan rich basic protein	0.000613
35223_at	AB023234	AIBP63	alpha integrin binding protein 63	0.000674
38711_at	AB014527	CLASP2	cytoplasmic linker associated protein 2	0.000792
1515_at	L37374	FEN1	flap structure-specific endonuclease 1; cleaves 5' FLAP DNA structures, XPG/RAD2 family	0.001349
36478_at	X83973	TTF1	transcription termination factor, RNA polymerase I	0.002113
UNDEREXPRESSED in translocations				
1992_at	U46922	FHIT	fragile histidine triad; nucleotide metabolism, tumor suppressor	0.000000
1971_g_at	U46922	FHIT	fragile histidine triad; nucleotide metabolism, tumor suppressor	0.000002
34904_at	S40369	GRIK5	Glutamate receptor, ionotropic, kainate 5; neurotransmission	0.000005
32786_at	X51345	JUNB	jun B proto-oncogene; sequence specific transcription factor	0.000010
34210_at	N90866	CDW52	CDW52 antigen (CAMPATH-1 antigen)	0.000033
1105_s_at	M12886	TRB	T cell receptor, beta	0.000073
33261_at	M16941	HLA-DRB1	major histocompatibility complex, class II, DR beta 1	0.000120

1007_s_at	U48705	DDR1	discoidin domain receptor family, member 1; tyrosine kinase, insulin receptor subfamily	0.000201
529_at	U15932	DUSP5	dual specificity phosphatase 5	0.000437
34491_at	AJ225089	OASL	2'-5'-oligoadenylate synthetase-like; binds RNA and DNA	0.000708
AFFX-M27830	M27830	28S rRNA	28S ribosomal RNA gene	0.001901
266_s_at	L33930	CD24	CD24 antigen (small cell lung carcinoma cluster 4 antigen); B-cell response	0.003808

Table 4| Translocation related Genes

OVEREXPRESSED in MLL compared to ALL and CD10-

Num	Accession	Symbol and name
1	M14087	HL14, beta-galactoside-binding lectin
2	L40377	SERPINB8, serine (or cystein) proteinase inhibitor, clade B (ovalbumin), member 8
3	AF072099	immunoglobulin-like transcript 3 protein, variant 1
4	M54992	CD72 antigen; B cell proliferation and differentiation
5	U41813	HOXA9, homeo box A9; sequence specific transcription factor, murine myeloid leukemia
6	M58597	FUT4, fucosyltransferase 4 (alpha (1,3) fucosyltransferase, myeloid-specific); glycosylation
7	AC004080	HOXA10, homeo box A10; sequence specific transcription factor
8	AA099265	RECK, reversion-inducing-cystein-rich protein with kazal motifs; membrane glycoprotein, tumor suppression
9	AF041248	CDKN2C, cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4); negative control of cell proliferation
10	D16532	VLDLR, very low density lipoprotein receptor; signal transduction, modulation of Dab1/Tau phosphorylation
11	D78177	QPRT, quinolinate phosphoribosyltransferase; biosynthesis of NAD and NADP
12	U85707	MEIS1, myeloid ecotropic viral integration site 1 homolog (mouse); homeobox protein, murine leukemia
13	Z69030	PPP2R5C, protein phosphatase 2, regulatory subunit B (B56), gamma isoform
14	AI535946	LGALS1, lectin, galactoside-binding, soluble 1 (galectin1); cell apoptosis and differentiation
15	L06419	PLOD, procollagen-lysine, 2-oxoglutarate 5-dioxygenase (lysine hydroxylase, Ehlers-Danlos syndrome)
16	V00568	MYC, v-myc myelocytomatosis viral oncogene homolog (avian); transcription factor, cell proliferation
17	AF070523	JWA, tumor rejection antigen (gp96) pseudogene 1
18	M59040	CD44, CD44 antigen (Indian blood group system); cell surface receptor, lymphocyte activation/homing
19	L05424	CD44, CD44 antigen (Indian blood group system); cell surface receptor, lymphocyte activation/homing

UNDEREXPRESSED in MLL compared to ALL and CD10-

20	L20433	POU4F1, POU domain class 4, transcription factor1; synaptogenesis, axonogenesis
21	S80864	CYCL, cytochrome c-like antigen; tumor antigen
22	AB007895	KIAA0435
23	AB002301	KIAA0303
24	AL050173	C21orf25, chromosome 21open reading frame 25
25	AF084481	WFS1, Wolfram syndrome 1 (wolframin)
26	U59912	MADH1, mothers against decapentaplegic homolog 1 (Drosophila); receptor-regulated transcription
27	Z22576	CD69, CD69 antigen (p60, early T-cell activation antigen); receptor, lymphocyte proliferation
28	M92357	TNFAIP2, tumor necrosis factor, alpha-induced protein 2
29	U70321	TNFRSF14, tumor necrosis factor receptor superfamily, member 14; lymphocyte activation
30	X76104	DAPK1, death-associated protein kinase 1; mediating interferon-gamma-induced cell death
31	X58072	GATA3, GATA-binding protein 3; transcriptional activator of T cell receptor alpha and delta genes
32	U19969	TCF8, transcription factor 8; repression of transcription (interleukin-2)
33	D13639	CCND2, cyclin D2; cell cycle control
34	U01062	ITPR3, inositol 1,4,5-triphosphate receptor type 3; signal transduction, small molecule transport
35	S80562	CNN3, calponin 3, acidic; thin filament-associated protein, smooth muscle contraction

36	X53586	ITGA6, integrin alpha 6; laminin receptor, critical structure role in the hemidesmosome
37	U46922	FHIT, fragile histidine triad; nucleotide metabolism, tumor suppressor
38	U46922	FHIT, fragile histidine triad; nucleotide metabolism, tumor suppressor
39	U01134	FLT1, fms-related tyrosine kinase 1, vascular endothelial growth factor receptor
40	J04111	JUN, v-jun sarcoma virus 17 oncogene homolog (avian); transcription factor
41	L34059	CDH4, cadherin 4, type 1, R-cadherin (retinal); cell adhesion
42	U59423	MADH1, mothers against decapentaplegic homolog 1 (Drosophila); receptor-regulated transcription
43	M16594	GSTA2, glutathione S-transferase A2
44	U03858	FLT3LG, fms-related tyrosine kinase 3 ligand; stimulates proliferation of early hematopoietic cells
45	S59184	RYK, RYK receptor-like tyrosine kinase
46	J03600	ALOX5, arachidonate 5- lipoxygenase; biosythesis of leukotrienes

Table 5| Genes associated with differentiation

OVEREXPRESSED in MLL +CD10- compared to ALL

Num	Accession	Symbol and name
1	X00351	ACTB, actin, beta
2	M11507	TFRC, transferrin receptor (p90, CD71)
3	Z48501	PABPC1, poly(A) binding protein, cytoplasmic 1
4	L22075	GNA13, guanine nucleotide binding protein (G protein), alpha 13
5	U28964	YWHAZ, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide
6	M17254	ERG, v-ets erythroblastosis virus E26 oncogene like (avian)
7	AF073362	MRE11A, MRE11 meiotic recombination 11 homolog A (S. cerevisiae)
8	AF073362	MRE11A, MRE11 meiotic recombination 11 homolog A (S. cerevisiae)
9	M91196	ICSBP1, interferon consensus sequence binding protein 1
10	L19161	EIF2S3, eukaryotic translation initiation factor 2, subunit 3 gamma, 52kDa
11	Z35102	STK38, serine/threonine kinase 38
12	X04409	GNAS, GNAS complex locus
13	X04409	GNAS, GNAS complex locus
14	AF026445	TAF2, TAF2 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 150kDa
15	AF038960	VPS4B, vacuolar protein sorting 4B (yeast)
16	L40411	TRIP8, thyroid hormone receptor interactor 8
17	M74091	CCNC, cyclin C
18	U40462	ZNFN1A1, zinc finger protein, subfamily 1A, 1 (Ikaros)
19	U63289	CUGBP1, CUG triplet repeat, RNA binding protein 1
20	L36140	RECQL, RecQ protein-like (DNA helicase Q1-like)
21	X16983	ITGA4, integrin, alpha 4 (antigen CD49D, alpha 4 subunit of VLA-4 receptor)
22	AF051323	SCAP2, src family associated phosphoprotein 2
23	D87459	WASF1, WAS protein family, member 1
24	X81889	PKP4, plakophilin 4
25	D14710	ATP5A1, ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1, cardiac muscle
26	AL080119	PAI-RBP1
27	AL080119	PAI-RBP1
28	Y00638	PTPRC, protein tyrosine phosphatase, receptor type, C
29	U10324	ILF3, interleukin enhancer binding factor 3, 90kDa
30	U12022	CALM2, calmodulin 2 (phosphorylase kinase, delta)
31	X64838	RSN, restin (Reed-Steinberg cell-expressed intermediate filament-associated protein)
32	D14043	CD164, CD164 antigen, sialomucin
33	AF098799	RANBP7, RAN binding protein 7
34	AF006082	ACTR2, ARP2 actin-related protein 2 homolog (yeast)
35	W28994	DNCI2, dynein, cytoplasmic, intermediate polypeptide 2

36	M21154	AMD1, S-adenosylmethionine decarboxylase 1 [
37	AF056490	PDE8A, phosphodiesterase 8A
38	X83928	TAF11, TAF11 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 28kDa
39	AF091083	LOC56270, hypothetical protein 628
40	H15872	H41, hypothetical protein H41
41	M86667	NAP1L1, nucleosome assembly protein 1-like 1
42	D26155	SMARCA2, SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2
43	X68836	MAT2A, methionine adenosyltransferase II, alpha
44	U34994	PRKDC, protein kinase, DNA-activated, catalytic polypeptide
45	X74594	RBL2, retinoblastoma-like 2 (p130)
46	L05624	MAP2K1, mitogen-activated protein kinase kinase 1
47	J04088	TOP2A, topoisomerase (DNA) II alpha 170kDa
48	L49229	RB1, retinoblastoma susceptibility protein
49	M27504	TOP2B, topoisomerase (DNA) II beta 180kDa
50	L12723	HSPA4, heat shock 70kDa protein 4
51	L19161	EIF2S3, eukaryotic translation initiation factor 2, subunit 3 gamma, 52kDa
52	L22075	GNA13, guanine nucleotide binding protein (G protein), alpha 13
53	U02687	FLT3, fms-related tyrosine kinase 3
54	U48251	PRKCBP1, protein kinase C binding protein 1
55	U48736	PRPF4B, PRP4 pre-mRNA processing factor 4 homolog B (yeast)
56	U50553	DDX3, DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 3
57	U77948	GTF2I, general transcription factor II, i
58	D28423	SRp20, pre-mRNA splicing factor
59	D30037	PITPNB, phosphotidylinositol transfer protein, beta
60	L13616	PTK2, PTK2 protein tyrosine kinase 2 (focal adhesion kinase 1)
UNDEREXPRESSED in MLL +CD10- compared to ALL		
61	D30758	CENTB1, centaurin, beta 1
62	M63928	TNFRSF7, tumor necrosis factor receptor superfamily, member 7
63	L16896	HKR3, GLI-Kruppel family member HKR3
64	U78190	GCHFR, GTP cyclohydrolase I feedback regulatory protein
65	AF035625	STK11, serine/threonine kinase 11 (Peutz-Jeghers syndrome)
66	AF060228	RARRES3, retinoic acid receptor responder (tazarotene induced) 3
67	AF002210	CCS, copper chaperone for superoxide dismutase
68	L25665	GNL1, guanine nucleotide binding protein-like 1
69	AF000424	LST1, leukocyte specific transcript 1
70	U83993	P2RX4, purinergic receptor P2X, ligand-gated ion channel, 4
71	AF001383	BIN1, bridging integrator 1
72	X91504	ARFRP1, ADP-ribosylation factor related protein 1
73	X90858	UP, uridine phosphorylase
74	Z46389	VASP, vasodilator-stimulated phosphoprotein
75	AB014585	KIAA0685
76	AF089750	FLOT1, flotillin 1
77	L25665	GNL1, guanine nucleotide binding protein-like 1
78	J04164	IFITM1, interferon induced transmembrane protein 1 (9-27)
79	U25265	MAP2K5, mitogen-activated protein kinase kinase 5
80	U68485	BIN1, bridging integrator 1

Table 6 | CD10- related genes

OVEREXPRESSED in CD10- compared with ALL and MLL

Num	Accession	Symbol and name
1	X59244	ZNF43, zinc finger protein 43 (HTF6)

2	U71364	SERPINB9, serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 9
3	U35376	ZNF85, zinc finger protein 85 (HPF4, HTF1)
4	W27883	SDCCAG1, serologically defined colon cancer antigen 1
5	AL030996	THO2
6	AB020683	KIAA0876, tumor rejection antigen (gp96) pseudogene 1
7	U77129	MAP4K5, mitogen-activated protein kinase kinase kinase kinase 5
8	U77413	OGT, O-linked N-acetylglucosamine (GlcNAc) transferase
9	S66213	ITGA6, integrin, alpha 6
10	S66213	ITGA6, integrin, alpha 6
11	X75755	SFRS2, splicing factor, arginine/serine-rich 2
12	X75042	REL, v-rel reticuloendotheliosis viral oncogene homolog (avian)
13	X06318	PRKCB1, protein kinase C, beta 1
14	X68277	DUSP1, dual specificity phosphatase 1
15	L11672	ZNF91, zinc finger protein 91 (HPF7, HTF10)
16	S53911	CD34, CD34 antigen
UNDEREXPRESSED in CD10- compared with ALL and MLL		
17	AI017574	CRIP1, cysteine-rich protein 1 (intestinal)
18	AF060228	RARRES3, retinoic acid receptor responder (tazarotene induced) 3
19	X71129	ETFB, electron-transfer-flavoprotein, beta polypeptide
20	U83993	P2RX4, purinergic receptor P2X, ligand-gated ion channel, 4
21	U81800	SLC16A3, solute carrier family 16 (monocarboxylic acid transporters), member 3

Table 7| Genes Found in the Two-Way ANOVA

Num	Gene ID	Accession	Symbol	Description	Function
1	1348_s_at	S79219	PCCA	propionyl Coenzyme A carboxylase, alpha polypeptide	GO:6631;fatty acid metabolism;not recorded
2	33263_at	X67098	HSRTSBETA	rTS beta protein	
3	33264_at	X89602	HSRTSBETA	rTS beta protein	
4	36498_at	AI936759	AP1S1	adaptor-related protein complex 1, sigma 1 subunit	GO:6899;non-selective vesicle transport;predicted/computed
5	1674_at	M15990	YES1	v-yes-1 Yamaguchi sarcoma viral oncogene homolog 1	GO:6464;protein modification;predicted/computed
6	32156_at	AF044968	PVRL2	poliovirus receptor-related 2 (herpesvirus entry mediator B)	GO:9406;virulence;experimental evidence
7	41266_at	X53586	ITGA6	integrin, alpha 6	GO:7160;cell-cell matrix adhesion;not recorded GO:7044;cell-substrate junction assembly;experimental evidence
8	37280_at	U59912	MADH1	MAD, mothers against decapentaplegic homolog 1 (Drosophila)	
9	33809_at	AL049933	GNAI1	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1	
10	40511_at	X58072	GATA3	GATA binding protein 3	GO:6366;transcription from Pol II promoter;experimental evidence GO:6952;defense response;predicted/computed GO:7345;embryogenesis and morphogenesis;experimental evidence
11	35164_at	AF084481	WFS1	Wolfram syndrome 1 (wolframin)	GO:6091;energy pathways;predicted/computed GO:7601;vision;predicted/computed GO:7399;neurogenesis;predicted/computed

12	1992_at	U46922	FHIT	fragile histidine triad gene	GO:9117;nucleotide metabolism;predicted/computed GO:7048;oncogenesis;predicted/computed
13	39569_at	U72849	EVPL	envoplakin	GO:8544;epidermal differentiation;predicted/computed
14	32107_at	AL050173	C21orf25	chromosome 21 open reading frame 25	
15	41656_at	AF043325	NMT2	N-myristoyltransferase 2	GO:9249;lipid:protein modification;experimental evidence
16	40953_at	S80562	CNN3	calponin 3, acidic	
17	35961_at	AL049390		Homo sapiens mRNA; cDNA DKFZp586O1318 (from clone DKFZp586O1318)	
18	33440_at	U19969	TCF8	transcription factor 8 (represses interleukin 2 expression)	GO:6357;transcription regulation from Pol II promoter;experimental evidence GO:6955;immune response;predicted/computed GO:122;repression of transcription from Pol II promoter;predicted/computed GO:8283;cell proliferation;predicted/computed
19	41710_at	AL079277	LOC54103	hypothetical protein LOC54103	
20	182_at	U01062	ITPR3	inositol 1,4,5-triphosphate receptor, type 3	GO:7165;signal transduction;experimental evidence GO:6832;small molecule transport;experimental evidence
21	31886_at	X55740	NT5E	5'-nucleotidase, ecto (CD73)	GO:6259;DNA metabolism;predicted/computed
22	37343_at	U01062	ITPR3	inositol 1,4,5-triphosphate receptor, type 3	GO:6832;small molecule transport;experimental evidence GO:7165;signal transduction;experimental evidence
23	39424_at	U70321	TNFRSF14	tumor necrosis factor receptor superfamily, member 14 (herpesvirus entry mediator)	GO:7166;cell surface receptor linked signal transduction;predicted/computed GO:6955;immune response;predicted/computed
24	37973_at	AB018256	SNX13	sorting nexin 13	
25	539_at	S59184	RYK	RYK receptor-like tyrosine kinase	GO:7165;signal transduction;predicted/computed
26	32554_s_at	Y12781	TBL1X	transducin (beta)-like 1X-linked	GO:7601;vision;predicted/computed GO:7605;hearing;predicted/computed GO:7165;signal transduction;predicted/computed
27	38833_at	X00457	HLA-DPA1	major histocompatibility complex, class II, DP alpha 1	
28	34376_at	AB019517	PKIG	protein kinase (cAMP-dependent, catalytic) inhibitor gamma	
29	37479_at	M54992	CD72	CD72 antigen	GO:6960;antimicrobial humoral response;not recorded GO:7155;cell adhesion;predicted/computed
30	1115_at	M25897	PF4	platelet factor 4 (chemokine (C-X-C motif) ligand 4)	
31	37009_at	AL035079	CAT	catalase	

32	33412_at	AI535946	LGALS1	lectin, galactoside-binding, soluble, 1 (galectin 1)	GO:6915;apoptosis;predicted/computed
33	41448_at	AC004080	HOXA10	homeo box A10	GO:7283;spermatogenesis;predicted/computed GO:7275;developmental processes;experimental evidence GO:7048;oncogenesis;experimental evidence
34	31575_f_at	M14087			
35	39210_at	M58597	FUT4	fucosyltransferase 4 (alpha (1,3) fucosyltransferase, myeloid-specific)	GO:5975;carbohydrate metabolism;experimental evidence
36	32184_at	X61118	LMO2	LIM domain only 2 (rhombotin-like 1)	GO:7275;developmental processes;predicted/computed GO:7048;oncogenesis;predicted/computed
37	32475_at	AF025529	LILRA1	leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 1	GO:7166;cell surface receptor linked signal transduction;predicted/computed GO:6952;defense response;predicted/computed
38	37724_at	V00568	MYC	v-myc myelocytomatosis viral oncogene homolog (avian)	GO:6357;transcription regulation from Pol II promoter;experimental evidence GO:8283;cell proliferation;experimental evidence GO:7050;cell cycle arrest;experimental evidence GO:9405;pathogenesis;predicted/computed GO:6879;iron homeostasis;experimental evidence
39	31472_s_at	AF098641		Homo sapiens CD44 isoform RC (CD44) mRNA, complete cds	
40	1126_s_at	L05424			
41	2036_s_at	M59040	CD44	CD44 antigen (homing function and Indian blood group system)	
42	39732_at	X73882	MAP7	microtubule-associated protein 7	GO:226;microtubule cytoskeleton organization and biogenesis;predicted/computed GO:7163;establishment of cell polarity;predicted/computed
43	37184_at	L37792	STX1A	syntaxin 1A (brain)	GO:6887;exocytosis;not recorded GO:6899;non-selective vesicle transport;not recorded
44	41470_at	AF027208	PROML1	prominin-like 1 (mouse)	
45	37251_s_at	AF016004	GPM6B	glycoprotein M6B	
46	38004_at	X96753	CSPG4	chondroitin sulfate proteoglycan 4 (melanoma-associated)	GO:7048;oncogenesis;not recorded GO:6928;cell motility;not recorded
47	1771_s_at	J03278	PDGFRB	platelet-derived growth factor receptor, beta polypeptide	GO:7165;signal transduction;experimental evidence
48	34961_at	M88282	TACTILE	T cell activation, increased late expression	GO:6955;immune response;predicted/computed GO:7155;cell adhesion;predicted/computed
49	31506_s_at	L12691			

Table 8| Genes of Cluster 1. Upregulated in AML with Translocation.

Num	Gene ID	Accession	Symbol	Description	Function
1	1348_s_at	S79219	PCCA	propionyl Coenzyme A	GO:6631;fatty acid metabolism;not recorded

				carboxylase, alpha polypeptide	
2	33263_at	X67098	HSRTSBETA	rTS beta protein	
3	33264_at	X89602	HSRTSBETA	rTS beta protein	
4	36498_at	AI936759	AP1S1	adaptor-related protein complex 1, sigma 1 subunit	GO:6899;non-selective vesicle transport;predicted/computed

Table 9| Genes of Cluster 2. Upregulated in ALL without Translocation.

Num	Gene ID	Accession	Symbol	Description	Function
6	32156_at	AF044968	PVRL2	poliovirus receptor-related 2 (herpesvirus entry mediator B)	GO:9406;virulence;experimental evidence
7	41266_at	X53586	ITGA6	integrin, alpha 6	GO:7160;cell-cell matrix adhesion;not recorded GO:7044;cell-substrate junction assembly;experimental evidence
8	37280_at	U59912	MADH1	MAD, mothers against decapentaplegic homolog 1 (Drosophila)	
9	33809_at	AL049933	GNAI1	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1	
10	40511_at	X58072	GATA3	GATA binding protein 3	GO:6366;transcription from Pol II promoter;experimental evidence GO:6952;defense response;predicted/computed GO:7345;embryogenesis and morphogenesis;experimental evidence
11	35164_at	AF084481	WFS1	Wolfram syndrome 1 (wolframin)	GO:6091;energy pathways;predicted/computed GO:7601;vision;predicted/computed GO:7399;neurogenesis;predicted/computed
12	1992_at	U46922	FHIT	fragile histidine triad gene	GO:9117;nucleotide metabolism;predicted/computed GO:7048;oncogenesis;predicted/computed
13	39569_at	U72849	EVPL	envoplakin	GO:8544;epidermal differentiation;predicted/computed
14	32107_at	AL050173	C21orf25	chromosome 21 open reading frame 25	

Table 10| Genes of Cluster 3. Upregulated in ALL with MLL Translocation.

Num	Gene ID	Accession	Symbol	Description	Function
43	37184_at	L37792	STX1A	syntaxin 1A (brain)	GO:6887;exocytosis;not recorded GO:6899;non-selective vesicle transport;not recorded
44	41470_at	AF027208	PROML1	prominin-like 1 (mouse)	
45	37251_s_at	AF016004	GPM6B	glycoprotein M6B	
46	38004_at	X96753	CSPG4	chondroitin sulfate proteoglycan 4 (melanoma-associated)	GO:7048;oncogenesis;not recorded GO:6928;cell motility;not recorded
47	1771_s_at	J03278	PDGFRB	platelet-derived growth factor receptor, beta polypeptide	GO:7165;signal transduction;experimental evidence

Table 11| Genes of Cluster 4. Downregulated in AMLs and ALLs with MLL translocation.

Num	Gene ID	Accession	Symbol	Description	Function
15	41656_at	AF043325	NMT2	N-myristoyltransferase 2	GO:9249;lipid:protein modification;experimental evidence
16	40953_at	S80562	CNN3	calponin 3, acidic	
17	35961_at	AL049390		Homo sapiens mRNA; cDNA DKFZp586O1318 (from clone DKFZp586O1318)	

18	33440_at	U19969	TCF8	transcription factor 8 (represses interleukin 2 expression)	GO:6357;transcription regulation from Pol II promoter;experimental evidence GO:6955;immune response;predicted/computed GO:122;repression of transcription from Pol II promoter;predicted/computed GO:8283;cell proliferation;predicted/computed
19	41710_at	AL079277	LOC54103	hypothetical protein LOC54103	
20	182_at	U01062	ITPR3	inositol 1,4,5-triphosphate receptor, type 3	GO:7165;signal transduction;experimental evidence GO:6832;small molecule transport;experimental evidence
21	31886_at	X55740	NT5E	5'-nucleotidase, ecto (CD73)	GO:6259;DNA metabolism;predicted/computed
22	37343_at	U01062	ITPR3	inositol 1,4,5-triphosphate receptor, type 3	GO:6832;small molecule transport;experimental evidence GO:7165;signal transduction;experimental evidence
23	39424_at	U70321	TNFRSF14	tumor necrosis factor receptor superfamily, member 14 (herpesvirus entry mediator)	GO:7166;cell surface receptor linked signal transduction;predicted/computed GO:6955;immune response;predicted/computed

Table 12| Genes of Cluster 5. Upregulated in both AMLs and in ALL with MLL translocation.

Num	Gene ID	Accession	Symbol	Description	Function
31	37009_at	AL035079	CAT	catalase	
32	33412_at	AI535946	LGALS1	lectin, galactoside-binding, soluble, 1 (galectin 1)	GO:6915;apoptosis;predicted/computed
33	41448_at	AC004080	HOXA10	homeo box A10	GO:7283;spermatogenesis;predicted/computed GO:7275;developmental processes;experimental evidence GO:7048;oncogenesis;experimental evidence
34	31575_f_at	M14087			
35	39210_at	M58597	FUT4	fucosyltransferase 4 (alpha (1,3) fucosyltransferase, myeloid-specific)	GO:5975;carbohydrate metabolism;experimental evidence
36	32184_at	X61118	LMO2	LIM domain only 2 (rhombotin-like 1)	GO:7275;developmental processes;predicted/computed GO:7048;oncogenesis;predicted/computed

Table 13| Genes of Cluster 6. Downregulated in ALL with no MLL translocation, upregulated in ALL with MLL translocation, and slightly upregulated in AML with no translocation.

Num	Gene ID	Accession	Symbol	Description	Function
38	37724_at	V00568	MYC	v-myc myelocytomatosis viral oncogene homolog (avian)	GO:6357;transcription regulation from Pol II promoter;experimental evidence GO:8283;cell proliferation;experimental evidence GO:7050;cell cycle arrest;experimental evidence

					GO:9405;pathogenesis;predicted/computed GO:6879;iron homeostasis;experimental evidence
39	31472_s_at	AF098641		Homo sapiens CD44 isoform RC (CD44) mRNA, complete cds	
40	1126_s_at	L05424			
41	2036_s_at	M59040	CD44	CD44 antigen (homing function and Indian blood group system)	
42	39732_at	X73882	MAP7	microtubule-associated protein 7	GO:226;microtubule cytoskeleton organization and biogenesis;predicted/computed GO:7163;establishment of cell polarity;predicted/computed

Table 14| Cluster 7, Separating ALL t(4:11) and CD10- from the Rest

Num	Gene ID	Accession	Symbol	Title and function
1	35484_at	U95737	FLJ00066	FLJ00066 protein
2	35939_s_at	L20433	POU4F1	POU domain, class 4, transcription factor 1; homeobox transcription factor, neuronal lineages
3	35940_at	X64624	POU4F1	POU domain, class 4, transcription factor 1; homeobox transcription factor, neuronal lineages
4	38567_at	L38820	CD1D	CD1D antigen, d polypeptide; immunoglobulin superfamily
5	39610_at	X16665	HOXB2	homeo box B2; sequence specific transcription factor
6	41385_at	AB023204	EPB41L3	erythrocyte membrane protein band 4.1-like 3; growth regulator of meningiomas
7	41439_at	AJ001381	MYO1B	myosin IB
8	41708_at	AB028957	KIAA1034	KIAA1034 protein; homeobox transcription factor, cut family
9	31886_at	X55740	NT5E	5'-nucleotidase, ecto (CD73); DNA metabolism, transport of nucleosides into cells
10	33244_at	U07223	CHN2	chimerin (chimaerin) 2; GTPase activator of P21-RAC
11	36821_at	AL050367	LOC221061	LOC221061
12	38323_at	AC005162	CPVL	carboxypeptidase, vitellogenic-like
13	40155_at	D31883	ABLIM	actin binding LIM protein
14	40425_at	M57730	EFNA1	ephrin-A1; induced by TNF, ephrin family
15	40504_at	AF001601	PON2	paraoxonase 2; hydrolysis of organophosphates and aromatic esters
16	41124_r_at	L35594	ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2 (autotaxin)
17	41188_at	W28186	LC27	putative integral membrane transporter
18	41209_at	M15856	LPL	lipoprotein lipase; fatty acid metabolism, hydrolysis of triglycerides
19	35766_at	M26326	KRT18	keratin 18; intermediate filament family
20	36617_at	X77956	ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein; transcription factor
21	37043_at	AL021154	ID3	inhibitor of DNA binding 3, dominant negative helix-loop-helix protein; tanscription factor
22	38077_at	X52022	COL6A3	collagen, type VI, alpha 3; extracellular matrix
23	40953_at	S80562	CNN3	calponin 3, acidic; regulation of muscle contraction
24	1389_at	J03779	MME	membrane metallo-endopeptidase (neutral endopeptidase, enkephalinase, CALLA, CD10)
25	424_s_at	X66945	FGFR1	fibroblast growth factor receptor 1 (fms-related tyrosine kinase 2, Pfeiffer syndrome)

Table 15| Cluster 8, Separating Cell Lines from the Primary Tumors

	Gene ID	Accession	Symbol	Title and function
1	31315_at	D84143	IGLJ3	immunoglobulin lambda joining 3
2	31439_f_at	X63095	RHCE	Rhesus blood group, CcEe antigens; integral plasma membrane protein
3	31522_f_at	Z80779	H2BFA	H2B histone family, member A
4	31523_f_at	Z80780	H2BFH	H2B histone family, member H
5	31524_f_at	Z80782	H2BFA	H2B histone family, member A

6	31525_s_at	J00153	HBA1	hemoglobin, alpha 1; oxygen transport
7	31528_f_at	Z83738	H2BFE	H2B histone family, member E
8	31586_f_at	X72475		rearranged immunoglobulin light variable kappa
9	31687_f_at	M25079	HBB	hemoglobin, beta; oxygen transport
10	31768_at	AL009179	H2AFC	H2A histone family, member C
11	31930_f_at	X63096	RHCE	Rhesus blood group, CcEe antigens; integral plasma membrane protein
12	31931_f_at	AI632247	RHCE	Rhesus blood group, CcEe antigens; integral plasma membrane protein
13	34105_f_at	AI147237	IGHG3	immunoglobulin heavy constant gamma 3 (G3m marker)
14	34157_f_at	AI200373	H2AFI	H2A histone family, member I
15	34627_at	X90763	KRTHA5	keratin, hair, acidic, 5
16	35127_at	AI039144	H2AFA	H2A histone family, member A
17	35530_f_at	X92997	IGLJ3	immunoglobulin lambda joining 3
18	35576_f_at	AL009179	H2BFC	H2B histone family, member C
19	35601_at	L00022	IGHE	immunoglobulin heavy constant epsilon
20	36347_f_at	AA873858	H2BFD	H2B histone family, member D
21	33499_s_at	AF067420	SNC73	SNC73 protein
22	33500_i_at	S71043		immunoglobulin heavy constant alpha 2
23	33501_r_at	S71043		immunoglobulin heavy constant alpha 2
24	33516_at	V00505	HBD	hemoglobin, delta
25	33993_at	M22919	MYL6	myosin, light polypeptide 6, alkali, smooth muscle and non-muscle
26	34490_f_at	AI189621	FSCN2	fascin homolog 2, actin-bundling protein, retinal (Strongylocentrotus purpuratus); actin cytoskeleton
27	34964_at	N35832	H3FB	H3 histone family, member B
28	36234_at	U79273	23933	23933 protein
29	36711_at	AL021977	MAFF	v-maf musculoaponeurotic fibrosarcoma oncogene homolog F (avian), transcription factor
30	36757_at	AL009179	H3FK	H3 histone family, member K
31	37127_at	AB023143	DEFCAP	death effector filament-forming Ced-4-like apoptosis protein
32	37165_f_at	X54534	RHCE	Rhesus blood group, CcEe antigens
33	37524_at	AB011421	STK17B	serine/threonine kinase 17b (apoptosis-inducing); apoptotic signaling
34	37864_s_at	Y14737	IGHG3	immunoglobulin heavy constant gamma 3 (G3m marker)
35	38194_s_at	M63438	IGKC	immunoglobulin kappa constant
36	38197_at	M64934	KEL	Kell blood group; integral plasma membrane protein, endopeptidase
37	38225_at	AF052728	KCNH2	potassium voltage-gated channel, subfamily H (eag-related), member 2
38	38499_s_at	D28113	MOBP	myelin-associated oligodendrocyte basic protein
39	38555_at	AB026436	DUSP10	dual specificity phosphatase 10; inactivation of MAP kinases
40	38585_at	M91036	HBG1	hemoglobin, gamma A; fetal hemoglobin F formation
41	38906_at	M61877	SPTA1	spectrin, alpha, erythrocytic 1 (elliptocytosis 2); actin filament organization, cytoskelet network
42	40296_at	AL023653	CXorf9	chromosome X open reading frame 9
43	40311_at	AF053356	TFR2	transferrin receptor 2; iron transport, peptidase family M28B
44	40647_at	Z32684	XK	Kell blood group precursor (McLeod phenotype); small molecule transport
45	41015_at	AB022017	PRKAA1	protein kinase, AMP-activated, alpha 1 catalytic subunit; fatty acid and cholesterol synthesis
46	41026_f_at	U05255	HEP2	glycophorin HeP2
47	31812_at	M24470	GMPR	guanosine monophosphate reductase; nucleobase, nucleoside, nucleotide and nucleic acid metabolism
48	32052_at	L48215	HBB	hemoglobin, beta; oxygen transport
49	32663_at	X64594	RHAG	Rhesus blood group-associated glycoprotein
50	33273_f_at	X57809	HSPA1A	heat shock 70kD protein 1A
51	33274_f_at	M18645	IGLJ3	immunoglobulin lambda joining 3

52	33336_at	M27819	SLC4A1	solute carrier family 4, anion exchanger, member 1 (erythrocyte membrane protein band 3, Diego blood group)
53	33705_at	L20971	PDE4B	phosphodiesterase 4B, cAMP-specific (phosphodiesterase E4 dunce homolog, Drosophila)
54	33759_at	X04327	BPGM	2,3-bisphosphoglycerate mutase; regulator hemoglobin oxygen affinity, carbohydrate metabolism
55	35224_at	AF070569	MGC14376	hypothetical protein MGC14376
56	36036_at	J05500	SPTB	spectrin, beta, erythrocytic (includes spherocytosis, clinical type I); actin filament organization, cytoskelet network
57	36037_g_at	J05500	SPTB	spectrin, beta, erythrocytic (includes spherocytosis, clinical type I); actin filament organization, cytoskelet network
58	36493_at	M33552	LSP1	lymphocyte-specific protein 1; actin cytoskeleton
59	36560_at	AB007950	KIAA0481	KIAA0481gene product
60	36829_at	AF022991	PER1	period homolog 1 (Drosophila); entrainment of circadian clock
61	36871_at	M60298	EPB42	erythrocyte membrane protein band 4.2; erythrocyte shape and mechanical properties control
62	37192_at	U28389	EPB49	erythrocyte membrane protein band 4.9 (dematin); actin-bundling protein
63	37285_at	X60364	ALAS2	aminolevulinate, delta-, synthase 2 (sideroblastic/hypochromic anemia); heme synthesis
64	37544_at	X64318	NFIL3	nuclear factor, interleukin 3 regulated
65	37649_at	M95623	HMBS	hydroxymethylbilane synthase
66	37970_at	AB028989	MAPK8IP3	mitogen-activated protein kinase 8 interacting protein 3
67	38361_g_at	AI688812	RASGRP2	RAS guanyl releasing protein 2 (calcium and DAG-regulated)
68	38674_at	AA115140	KIAA1354	KIAA1354 protein
69	38714_at	L31860	GYPA	glycophorin A (includes MN blood group); major intrinsic membrane protein of the erythrocyte
70	38715_at	W04490	GYPB	glycophorin B (includes Ss blood group); minor sialoglycoprotein in erythrocyte membranes
71	39729_at	L19185	PRDX2	peroxiredoxin 2; reduction and elimination peroxides, enhances natural killer cell activity
72	39736_at	M35543	CDC42	cell division cycle 42 (GTP binding protein, 25kD); filopodia formation, GTPase superfamily, Rho family
73	40062_s_at	X58851	MYL4	myosin, light polypeptide 4, alkali; atrial, embryonic
74	40095_at	J03037	CA2	carbonic anhydrase II
75	40448_at	M92843	ZFP36	zinc finger protein 36, C3H type, homolog (mouse); mRNA catabolism, transcription factor
76	40814_at	L40586	IDS	iduronate 2-sulfatase (Hunter syndrome); degradation of heparin sulfate, dermatin sulfate
77	32189_g_at	M96980	MYT1	myelin transcription factor 1; neurogenesis
78	32786_at	X51345	JUNB	jun B proto-oncogene; DNA-binding transcription factor
79	32819_at	AJ223352	H2B/S	H2B histone family member
80	33352_at	X57985	H2BFQ	H2B histone family, member Q
81	33439_at	D15050	TCF8	transcription factor 8 (represses interleukin 2 expression); immune response
82	33834_at	L36033	SDF1	stromal cell-derived factor 1; chemoattractant on T-lymphocites, monocytes, signal transduction
83	34308_at	U90551	H2AFL	H2A histone family, member L
84	35785_at	W28281	GABARAPL1	GABA(A) receptor-associated protein like 1
85	35851_g_at	AI950382	PSR	phosphatidylserine receptor
86	36118_at	AJ000882	NCOA1	nuclear receptor coactivator 1; transcription co-activator
87	36575_at	S59049	RGS1	regulator of G-protein signalling 1; regulation B-cell activation and proliferation
88	36629_at	AI635895	DSIPI	delta sleep inducing peptide, immunoreactor
89	36630_at	Z50781	DSIPI	delta sleep inducing peptide, immunoreactor
90	36636_at	M12267	OAT	ornithine aminotransferase (gyrate atrophy); pyridoxal-phosphate dependent, aminoacid metabolism
91	36669_at	L49169	FOSB	FBJ murine osteosarcoma viral oncogene homolog B; JUN proteins interactor, transcription factor
92	36979_at	M20681	SLC2A3	solute carrier family 2 (facilitated glucose transporter), member 3; carbohydrate metabolism
93	36980_at	U03105	PROL2	proline rich 2

94	37002_at	D32143	BLVRB	biliverdin reductase B (flavin reductase (NADPH));electron transfer, iron metabolism, oxidative damage protection
95	37018_at	AI189287	H1F2	H1 histone family, member 2
96	37026_at	AF001461	COPEB	core promoter element binding protein; transcription factor
97	37028_at	U83981	PPP1R15A	protein phosphatase 1, regulatory (inhibitor) subunit 15A; cell cycle arrest, apoptosis
98	37294_at	X61123	BTG1	B-cell translocation gene 1, anti-proliferative; cell cycle regulator,
99	37405_at	U29091	SELENBP1	selenium binding protein 1
100	38393_at	D87434	KIAA0247	KIAA0247 gene product
101	38767_at	AF041037	SPRY1	sprouty homolog 1, antagonist of FGF signaling (Drosophila); fibroblast growth factor pathway, cell-cell signaling
102	38823_s_at	AI961743	STK17A	serine/threonine kinase 17a (apoptosis-inducing); phosphorylates myosin light chain, induction of apoptosis
103	39908_at	AF069735	TAF6L	TAF6-like RNA polymerase II, p300/CBP-associated factor (PCAF)-associated factor, 65 kD, HDAC complex, chromatin modelling
104	40227_at	D29810	ESDN	endothelial and smooth muscle cell-derived neuropilin-like protein
105	40926_at	U36341	SLC6A8	solute carrier family 6 (neurotransmitter transporter, creatine), member 8; creatine transporter
106	41544_at	AF059617	SNK	serum-inducible kinase; emryogenesis
107	41827_f_at	AI932613	FLJ32313	FLJ32313
108	32583_at	J04111	JUN	v-jun sarcoma virus 17 oncogene homolog (avian); DNA binding transcription factor, interacts with FOS, SMAD3/SMAD4
109	32588_s_at	X78992	ZFP36L2	zinc finger protein 36, C3H type-like 2; cell proliferation
110	32609_at	AI885852	H2AFO	H2A histone family, member O
111	2094_s_at	K00650	FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog; transcription factor, signal transduction, cell proliferatin, differentiation
112	2049_s_at	M29039	JUNB	jun B proto-oncogene; transcription factor, primary growth factor response
113	1915_s_at	V01512	FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog; transcription factor, signal transduction, cell proliferatin, differentiation
114	1916_s_at	V01512	FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog; transcription factor, signal transduction, cell proliferatin, differentiation
115	1891_at	D14497	MAP3K8	mitogen-activated protein kinase kinase kinase 8; activates NF-kappa-B 1, cell cycle
116	1895_at	J04111	JUN	v-jun sarcoma virus 17 oncogene homolog (avian); DNA binding transcription factor, interacts with FOS, SMAD3/SMAD4
117	1461_at	M69043	NFKBIA	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha; cell adhesion, immune response, apoptosis
118	1251_g_at	M64788	RAP1GA1	RAP1, GTPase activating protein 1; signal transduction
119	1237_at	S81914	IER3	immediate early response 3; apoptosis inhibitor, cell growth, embryogenesis
120	1115_at	M25897	PF4	platelet factor 4
121	1005_at	X68277	DUSP1	dual specificity phosphatase 1; oxidative stress and heat shock response
122	680_s_at	J05253	RBP3	retinol binding protein 3, interstitial; fatty acid binding, vision, interphotoreceptor extracellular matrix
123	669_s_at	L05072	IRF1	interferon regulatory factor 1; transcription activating factor
124	595_at	M59465	TNFAIP3	tumor necrosis factor, alpha-induced protein 3; apoptosis
125	560_s_at	M63589	TAL1	T-cell acute lymphocytic leukemia 1; binds to rhombotin-2, erythroid differentiation
126	529_at	U15932	DUSP5	dual specificity phosphatase 5
127	293_at	X74864	HPX42	homeo box protein
128	286_at	L19779	H2AFO	H2A histone family, member O
129	287_at	L19871	ATF3	activating transcription factor 3; binds to camp response element (cre) consensus
130	245_at	M25280	SELL	selectin L (lymphocyte adhesion molecule 1); cell motility, cell adhesion, integral plasma membrane protein
131	137_at	U65404	KLF1	Kruppel-like factor 1 (erythroid); sequence specific transcription factor, erythroid development
132	153_f_at	X00088	H2BFR	H2B histone family, member R

Table 16 | Patient Data

type	patient/cell line	sample	age	sex
T-cell ALL	Patient	ht1	41	m
T-cell ALL	Patient	ht2	15	m
T-cell ALL	Patient	ht3	48	m
CD10+ALL	Patient	ht5	24	m
CD10+ALL	Patient	ht6	50	m
CD10+ALL	Patient	ht7	24	f
Ph+ALL	Patient	ht9	41	f
Ph+ALL	Patient	ht10	56	f
Ph+ALL	Patient	ht11	57	m
CD10-ALL	Patient	ht12	35	m
t(4;11)ALL	Patient	ht13	24	m
CD10-ALL	Patient	ht14	24	m
CD10-ALL	Patient	ht15	15	f
CD10-ALL	Patient	ht16	18	m
t(4;11)ALL	Patient	ht17	19	f
t(4;11)ALL	Patient	ht18	39	f
t(4;11)ALL	Patient	ht19	36	m
t(4;11)ALL	Patient	ht20	58	f
t(4;11)ALL	Patient	ht21	15	m
t(4;11)ALL	Patient	ht22	0.8	m
t(4;11)ALL	Patient	ht23	40	f
t(4;11)ALL	Patient	ht24	42	f
t(4;11)ALL	Patient	ht25	57	m
t(4;11)ALL	Patient	ht26	38	f
t(4;11)ALL	Patient	ht27	52	f
t(4;11)ALL	B1-cell line	ht27.1		
t(4;11)ALL	RS(4;11)-cell line	ht27.2		
t(9;11)M5AML	Patient	ht34	38	m
t(9;11)M5AML	Patient	ht35		
t(9;11)M5AML	Patient	ht36	53	f
t(9;11)M5AML	Patient	ht37	60	f
t(9;11)M5AML	Patient	ht38	54	f
t(9;11)AML	TPH1-cell line	ht38.1		
t(9;11)AML	MONO6-cell line	ht38.2		
t(9;11)AML	MOL13-cell line	ht38.3		
t(9;11)AML	PER377-cell line	ht38.4		
MLLdup.M4AML	Patient	ht39	58	m
MLLdup.M5bAML	Patient	ht40	54	f
MLLdup.M5aAML	Patient	ht41	55	f
t(10;11)M5AML	Patient	ht42	45	f
t(11;19)AML	Patient	ht43	51	f
t(6;11)AML	ML2-cell line	ht44		
M4AML	Patient	ht45	48	f
M4AML	Patient	ht46	37	m
M4AML	Patient	ht47	54	f
M5AML	Patient	ht48	52	m
M5AML	Patient	ht48.1	58	f

M5AML	Patient	ht48.2	48	f
M5aAML	Patient	ht49	49	f
M5aAML	Patient	ht50	21	f
M5aAML	Patient	ht51	24	f
M5aAML	Patient	ht52	39	f

Reference List

Armstrong,S.A., Staunton,J.E., Silverman,L.B., Pieters,R., den Boer,M.L., Minden,M.D., Sallan,S.E., Lander,E.S., Golub,T.R., and Korsmeyer,S.J. (2002). MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat. Genet. *30*, 41-47.

Rozovskaia,T., Feinstein,E., Mor,O., Foa,R., Blechman,J., Nakamura,T., Croce,C.M., Cimino,G., and Canaani,E. (2001). Upregulation of Meis1 and HoxA9 in acute lymphocytic leukemias with the t(4 : 11) abnormality. Oncogene *20*, 874-878.

Yeoh,E.J., Ross,M.E., Shurtleff,S.A., Williams,W.K., Patel,D., Mahfouz,R., Behm,F.G., Raimondi,S.C., Relling,M.V., Patel,A., Cheng,C., Campana,D., Wilkins,D., Zhou,X., Li,J., Liu,H., Pui,C.H., Evans,W.E., Naeve,C., Wong,L., and Downing,J.R. (2002). Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell *1*, 133-143.

Appendix B

In addition to the 46 genes found to be in the intersection between the results of the comparisons MLL vs. ALL and MLL vs. CD10-, there were two additional genes (Affymetrix IDs 33236_at, 38332_at) which also came out in the comparison but were left out. These genes separated MLL from the other groups but in opposite directions.
Thus, these genes were high in the pre-B and T-cell ALL samples, intermediate in the MLL samples and low on the CD10- (Figure 1).

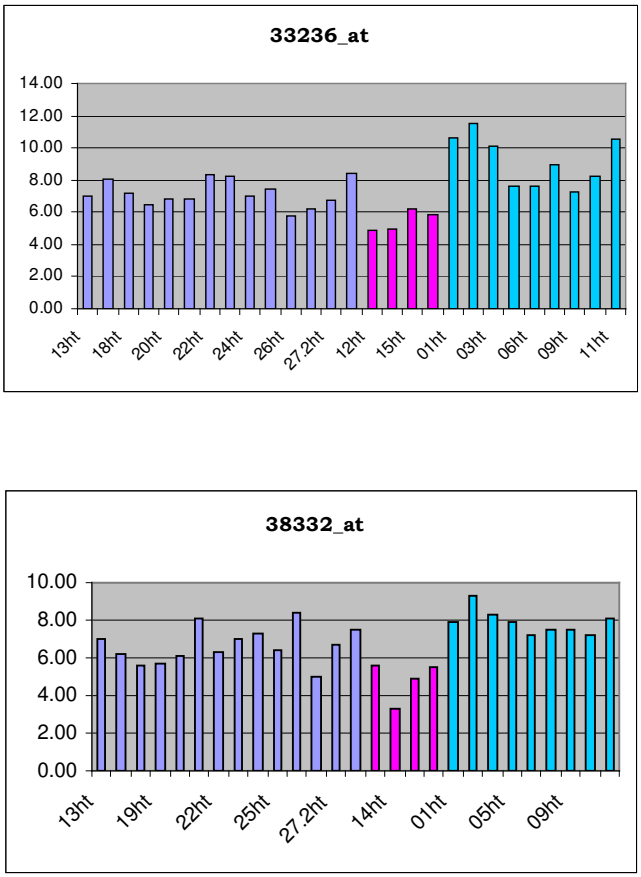

Figure 1 | **Genes separating MLL from the other groups but in different directions**. MLL in purple, CD10- in pink and other ALLs in turquoise. Gene separating MLL from CD10- and ALL.

In a similar way, there was an additional gene (Affymetrix ID 39424_at) appearing in the triple intersection. This gene was not

included in the triple intersection since it separated CD10- samples from ALL and MLL in opposite directions exhibiting a low level in MLL, high levels in CD10- and slightly higher in the pre-B and in T-cell ALLs (Figure 2). This gene was classified by us according to its values to the translocation related genes and was not included in the differentiation genes and the CD10- genes.

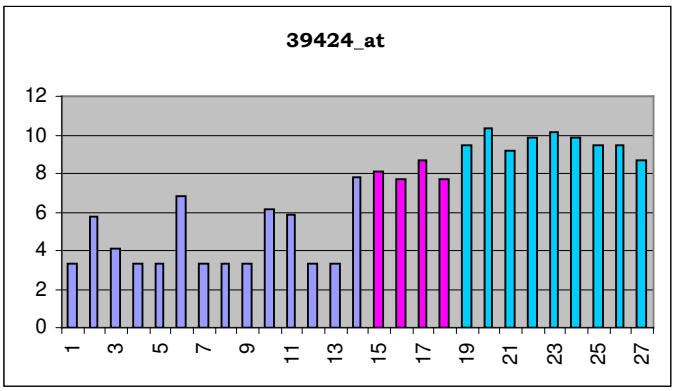

Figure 2 | Gene separating all three groups. MLL in purple, CD10- in pink and other ALLs in turquoise. Gene separating MLL from CD10- and ALL.

Appendix C

Table 1| Genes of Cluster 1. "Immunoglobulin cluster"

Probe Set ID	Accession	Symbol	Description
92470_f_at	AF065324	AI893585	expressed sequence AI893585
102823_at	X67210	AU044919	expressed sequence AU044919
101870_at	V00793	Igh-4	immunoglobulin heavy chain 4 (serum IgG1)
101871_f_at	X88902	Igh-4	immunoglobulin heavy chain 4 (serum IgG1)
101745_f_at	X94422	Igh-6	immunoglobulin heavy chain 6 (heavy chain of IgM)
100299_f_at	U55576	Igk-V28	immunoglobulin kappa chain variable 28 (V28)
101331_f_at	U55641	Igk-V28	immunoglobulin kappa chain variable 28 (V28)
93213_at	AB007986	Igk-V28	immunoglobulin kappa chain variable 28 (V28)
101640_f_at	M15593	Igk-V28	
101718_f_at	U19315	Igk-V28	
93227_at		Igk-V28	
100682_f_at	U62386		Mus musculus immunoglobulin heavy and light chain variable region
102843_s_at	J00475		Mus musculus, Similar to immunoglobulin heavy chain 1 (serum IgG2a)
93927_f_at	L33954		Mus musculus, Similar to immunoglobulin heavy chain 1 (serum IgG2a)
100721_f_at	U10410	LOC56304	recombinant antineuraminidase single chain Ig VH and VL domains
111469_at	AA794875	2010001M09Rik	RIKEN cDNA 2010001M09 gene
100360_f_at	X02466		
100362_f_at	X02463		
100376_f_at	AF025445		
100377_f_at	L33952		
101319_f_at	L28060		
101320_f_at	L28059		
101743_f_at	X94420		
101747_f_at	X94419		
101751_f_at	L43568		
101752_f_at	Z22111		
102076_at	AJ235940		
102155_f_at	K03461		
102157_f_at	M15520		
93904_f_at	AF059706		
96964_at	L14554		
96970_at	X16740		
96972_f_at	X00651		
96973_f_at	X02468		
97008_f_at	L33943		
97009_f_at	L33937		
97563_f_at	AF042798		
97566_f_at	AF045024		
97567_f_at	AF045026		
97574_f_at	AF036736		
97575_f_at	AF036737		
97576_f_at	AF036738		
97577_f_at	AF042086		
99369_f_at	AF029261		

Table 2 | Genes of Cluster 3. Collagens, Proteolytic Enzymes, and WNT.

Cell Adhesion,Collagen, and Extracellular Matrix

Probe ID	Accession	Symbol	Description
100308_at	X66976	Col8a1	procollagen, type VIII, alpha 1
102070_at	AW212495	Col9a3	procollagen, type IX, alpha 3
103616_at	AF100956	Col11a2	procollagen, type XI, alpha 2
104483_at	L12215	Col9a1	procollagen, type IX, alpha 1
100476_at	L20232	Ibsp	integrin binding sialoprotein
100481_at	D38162	Col11a1	procollagen, type XI, alpha 1
102261_f_at	U30292	Col13a1	procollagen, type XIII, alpha 1
163036_at	AV243867	Col9a1	procollagen, type IX, alpha 1
163627_f_at	AI847474	Cthrc1	collagen triple helix repeat containing 1
92207_at	U08210	Eln	Elastin
100336_s_at	L24431	Bglap2	bone gamma-carboxyglutamate protein 2
99903_at	AJ242625	Dmp1	dentin matrix protein 1
93369_at	AB007848	Omd	osteomodulin
116382_at	AW046112	Nrp2	neuropilin 2

Proteoglycans

Probe ID	Accession	Symbol	Description
93136_at	D78274	Dspg3	dermatan sulphate proteoglycan 3
104205_at	L07049	Agc1	aggrecan 1
160561_at	M34094	Mdk	Midkine

Wnt pathway related

Probe ID	Accession	Symbol	Description
103513_at	M89799	Wnt5b	wingless-related MMTV integration site 5B
93188_at	AJ243964	Dkk3	dickkopf homolog 3 (Xenopus laevis)
106051_at	AI838192	Dkk3	dickkopf homolog 3 (Xenopus laevis)
104672_at	U68058	Frzb	frizzled-related protein
163423_at	AW107799	Wif1	Wnt inhibitory factor 1
166263_at	AV346158	Frzb	frizzled-related protein
116807_at	AI846342	Mfrp	membrane-type frizzled-related protein

Proteases

Probe ID	Accession	Symbol	Description
97941_at	Y12582	Capn6	calpain 6
97942_g_at	Y12582	Capn6	calpain 6
99958_at	J05177	Mcpt2	mast cell protease 2
97109_at	L00653	Mcpt7	mast cell protease 7
100484_at	X66473	Mmp13	matrix metalloproteinase 13
99957_at	X72795	Mmp9	matrix metalloproteinase 9
165565_r_at	AI481262	Pcolce2	procollagen C-endopeptidase enhancer 2
108365_at	AA981778	Pcolce2	procollagen C-endopeptidase enhancer 2

Ligands (Growth factors), Cell growth

Probe ID	Accession	Symbol	Description
103244_at	AB009245	Scgf	stem cell growth factor
108045_at	AA798520	Lepre1	leprecan 1
165981_i_at	AV140849	Lect1	leukocyte cell derived chemotaxin 1
114346_at	AA960023	Il17b	interleukin 17B

Bone development

Probe ID	Accession	Symbol	Description
98482_at	X78936	Pthr	parathyroid hormone receptor
92796_at	J02980	Akp2	alkaline phosphatase 2, liver
162531_at	AW048375	Bambi	BMP and activin membrane-bound inhibitor, homolog (Xenopus laevis)

Transcription factors

Probe ID	Accession	Symbol	Description
133900_at	AI481773	Runx1	runt related transcription factor 1
135401_at	AW214326		ESTs, Highly similar to SOX8_MOUSE Transcription factor SOX-8 [M.musulus]

Others

Probe ID	Accession	Symbol	Description
100089_at	M74227	Ppic	peptidylprolyl isomerase C
103570_at	AI315647		Mus musculus, Similar to C1q and tumor necrosis factor related protein 3, clone IMAGE:3989958, mRNA
104477_at	AW047643		ESTs
92607_at	AF017994	Mest	mesoderm specific transcript
94181_at	AJ223206	Scrg1	scrapie responsive gene 1
99010_at	AB024538	Islr	immunoglobulin superfamily containing leucine-rich repeat
104932_at	AA197852	AI505012	expressed sequence AI505012
107935_at	AI450518	1110039C07 Rik	RIKEN cDNA 1110039C07 gene
107988_at	AI849414		Mus musculus, clone MGC:40948 IMAGE:5376413, mRNA, complete cds
108018_at	AA959436	2810002E22 Rik	RIKEN cDNA 2810002E22 gene
109033_at	AW122053	MGC27616	hypothetical protein MGC27616
112245_at	AW122302		ESTs, Weakly similar to growth factor receptor bound protein 14 [Mus musculus] [M.musulus]
113196_at	AW124306	MGC38046	hypothetical protein MGC38046
113851_at	AA204063	4631426B19 Rik	RIKEN cDNA 4631426B19 gene
115210_at	AA985762	5730512J02 Rik	RIKEN cDNA 5730512J02 gene
116181_at	AI020226	5730512J02 Rik	RIKEN cDNA 5730512J02 gene
117179_at	AI852894		ESTs, Weakly similar to JQ1322 tenascin precursor - mouse [M.musulus]
163213_at	AI648953		ESTs, Weakly similar to growth factor receptor bound protein 14 [Mus musculus] [M.musulus]
163245_at	AI155549		Mus musculus, Similar to DKFZP586L151 protein, clone MGC:38296 IMAGE:5342879, mRNA, complete cds
163765_at	AA388951	BC010311	cDNA sequence BC010311
132364_I_at	AI594430		ESTs
132365_r_at	AI594430		ESTs
101467_at	L22144	AI850290	expressed sequence AI850290
160971_at	AI842353	AI842353	expressed sequence AI842353
96481_at	AV251613	C80638	expressed sequence C80638
96835_at	AJ010984	Matn4	matrilin 4
113640_at	AW045201	1110004G24 Rik	RIKEN cDNA 1110004G24 gene
114224_at	AI425989		ESTs
115926_at	AA839289	5133401H06 Rik	RIKEN cDNA 5133401H06 gene
115962_at	AI615325		ESTs
163817_at	AI591564	1200015L10 Rik	RIKEN cDNA 1200015L10 gene

		Rik	
163903_at	AW124651		Mus musculus, clone IMAGE:4983969, mRNA
131792_s_at	AI850822		ESTs, Weakly similar to ENP1_MOUSE Ectonucleosidetriphosphate diphosphohydrolase 1 (NTPDase1) (Ecto-ATP diphosphohydrolase) (ATPDase) (Lymphoid cell activation antigen) (CD39 antigen) (Ecto-apyrase) [M.musulus]
134747_at	AI662497	8430438D04 Rik	RIKEN cDNA 8430438D04 gene
135131_at	AI508877		ESTs, Weakly similar to ITA2_MOUSE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GPIa) (Collagen receptor) (VLA-2 alpha chain) (CD49b) [M.musulus]
100472_at	D10727	Enah	enabled homolog (Drosophila)
165432_i_at	AI854199	Ubce8	ubiquitin-conjugating enzyme 8
102352_at	AW121380	Galnt9	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 9
96713_at	AF052453	Papss2	3'-phosphoadenosine 5'-phosphosulfate synthase 2

Table 3 | Complete List of Genes of cluster 4

Angiogenesis and ligands (growth factors) and their receptors

Gene ID	Symbol	Description
M32490_at	Cyr61	cysteine rich protein 61
M32745_at	Tgfb3	transforming growth factor, beta 3
X04367_at	Pdgfrb	platelet derived growth factor receptor, beta polypeptide
AF056187_at	Igf1r	insulin-like growth factor I receptor
M70642_at	Ctgf	connective tissue growth factor
M13177_at	Tgfb1	transforming growth factor, beta 1
L08394_at	Btc	betacellulin, epidermal growth factor family member

Cell cycle and apoptosis

Gene ID	Symbol	Description
Z26580_at	Ccna2	cyclin A2
Y13087_at	Casp6	caspase 6
U24173_at	Cdkn1a	cyclin-dependent kinase inhibitor 1A (P21)

WNT

Gene ID	Symbol	Description
AF139183_at	Fzd2	frizzled homolog 2 (Drosophila)
U43321_at	FZD8	frizzled homolog 8 (Drosophila)
AF126063_at	Wisp2	WNT1 inducible signaling pathway protein 2

Collagens, Extracellular Matrix

Gene ID	Symbol	Description
M18933_at	Col3a1	procollagen, type III, alpha 1
AB009993_at	Col5a1	procollagen, type V, alpha 1
M65161_x_at	Col2a1	procollagen, type II, alpha 1
U25652_at	Col12a1	procollagen, type XII, alpha 1
L02918_at	Col5a2	procollagen, type V, alpha 2
D17511_at	Col9a1	procollagen, type IX, alpha 1

AF080572_x_at	Plod2	procollagen lysine, 2-oxoglutarate 5-dioxygenase 2
M65142_x_at	Lox	lysyl oxidase

Proteoglycans

Gene ID	Symbol	Description
X15487_x_at	Sdc1	syndecan 1
D89571_at	Sdc4	syndecan 4
X70296_at	Serpine2	serine (or cysteine) proteinase inhibitor, clade E, member 2
X53928_at	Bgn	biglycan

Proteases

Gene ID	Symbol	Description
U54984_at	Mmp14	matrix metalloproteinase 14 (membrane-inserted)
M84324_at	Mmp2	matrix metalloproteinase 2
Z12604_at	Mmp11	matrix metalloproteinase 11
AB008548_at	Pcolce	procollagen C-proteinase enhancer protein

Adhesion

Gene ID	Symbol	Description
AF112151_at	Fbln5	fibulin 5
M27129_x_at	Cd44	CD44 antigen
X07233_at	Ncam1	neural cell adhesion molecule 1
AB008811_x_at	Cdh2	cadherin 2
X61576_at	Gja1	gap junction membrane channel protein alpha 1
D21253_at	Cdh11	cadherin 11
X56304_at	Tnc	tenascin C

BMP-related

Gene ID	Symbol	Description
AF010133_at	Madh6	MAD homolog 6 (Drosophila)
Z29532_at	Fst	follistatin
L25602_at	Bmp2	bone morphogenetic protein 2
L24755_at	Bmp1	bone morphogenetic protein 1
U04673_at	Bmpr1a	bone morphogenetic protein receptor, type 1A

Transcription Factors

Gene ID	Symbol	Description
M98339_at	Gata4	GATA binding protein 4
D14636_at	Runx2	runt related transcription factor 2

Activins, follistatin

Gene ID	Symbol	Description
X69619_x_at	Inhba	inhibin beta-A
Z29532_at	Fst	follistatin
X69620_at	Inhbb	inhibin beta-B

Inflammation

Gene ID	Symbol	Description
M72394_at	Pla2g4a	phospholipase A2, group IVA (cytosolic, calcium-dependent)
M64291_at	Ptgs2	prostaglandin-endoperoxide synthase 2

Others

Gene ID	Symbol	Description
X80638_x_at	Arhc	ras homolog gene family, member C
Z71173_x_at	Itpr5	inositol 1,4,5-triphosphate receptor 5
X54511_x_at	Capg	capping protein (actin filament), gelsolin-like
L08115_at	Cd9	CD9 antigen
X16074_at	Lgals3	lectin, galactose binding, soluble 3
M17243_at	TIMP	
K02927_at	Rrm1	ribonucleotide reductase M1
J03880_at	B4galt1	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 1
TYN1_x_at	TYN1	
AJ002390_at	Anxa8	annexin A8
X62700_at	Plaur	urokinase plasminogen activator receptor
L08394_at	Btc	betacellulin, epidermal growth factor family member
U13921_at	Krt1-13	keratin complex 1, acidic, gene 13
U79550_at	SLUGH	
M31885_x_at	Idb1	inhibitor of DNA binding 1
L26316_at	Dhfr	dihydrofolate reductase
Z49086_at	Ephb3	

Table 4 | Ordered List of Genes of Cluster 5. Separates Bas and Exp, chronologically orders the Exp samples

Num	Probe ID	Accession	Symbol	Description
1	112245_at	AW122302		ESTs, Weakly similar to growth factor receptor bound protein 14 [Mus musculus] [M.musulus]
2	104932_at	AA197852	AI505012	expressed sequence AI505012
3	101918_at	AJ009862	Tgfb1	transforming growth factor, beta 1
4	117179_at	AI852894		ESTs, Weakly similar to JQ1322 tenascin precursor - mouse [M.musulus]
5	97941_at	Y12582	Capn6	calpain 6
6	97942_g_at	Y12582	Capn6	calpain 6
7	92607_at	AF017994	Mest	mesoderm specific transcript
8	108045_at	AA798520	Lepre1	leprecan 1
9	99010_at	AB024538	Islr	immunoglobulin superfamily containing leucine-rich repeat
10	108365_at	AA981778	Pcolce2	procollagen C-endopeptidase enhancer 2
11	103570_at	AI315647		Mus musculus, Similar to C1q and tumor necrosis factor related protein 3, clone IMAGE:3989958, mRNA
12	107935_at	AI450518	1110039C07Rik	RIKEN cDNA 1110039C07 gene
13	104205_at	L07049	Agc1	aggrecan 1
14	104672_at	U68058	Frzb	frizzled-related protein
15	**92506_at**	AF098460	**Crtl1**	**cartilage link protein 1**
16	94181_at	AJ223206	Scrg1	scrapie responsive gene 1
17	116181_at	AI020226	5730512J02Rik	RIKEN cDNA 5730512J02 gene
18	**116582_at**	AI987726	**AW553532**	**expressed sequence AW553532**
19	114346_at	AA960023	Il17b	interleukin 17B
20	96835_at	AJ010984	Matn4	matrilin 4
21	100481_at	D38162	Col11a1	procollagen, type XI, alpha 1
22	101467_at	L22144	AI850290	expressed sequence AI850290
23	114224_at	AI425989		ESTs
24	133900_at	AI481773	Runx1	runt related transcription factor 1

25	135401_at	AW214326		ESTs, Highly similar to SOX8_MOUSE Transcription factor SOX-8 [M.musulus]
26	116807_at	AI846342	Mfrp	membrane-type frizzled-related protein
27	108018_at	AA959436	2810002E22Rik	RIKEN cDNA 2810002E22 gene
28	**114420_at**	AA734866	**Pcsk6**	**proprotein convertase subtilisin/kexin type 6**
29	**116956_at**	AI848366	**AU046135**	**expressed sequence AU046135**
30	103244_at	AB009245	Scgf	stem cell growth factor
31	100308_at	X66976	Col8a1	procollagen, type VIII, alpha 1
32	115210_at	AA985762	5730512J02Rik	RIKEN cDNA 5730512J02 gene
33	113196_at	AW124306	MGC38046	hypothetical protein MGC38046
34	**93485_at**	AI844911	**3000002J10Rik**	**RIKEN cDNA 3000002J10 gene**
35	96481_at	AV251613	C80638	expressed sequence C80638
36	116382_at	AW046112	Nrp2	neuropilin 2
37	**93351_at**	U44389	**Hpgd**	**hydroxyprostaglandin dehydrogenase 15 (NAD)**
38	106051_at	AI838192	Dkk3	dickkopf homolog 3 (Xenopus laevis)
39	115926_at	AA839289	5133401H06Rik	RIKEN cDNA 5133401H06 gene
40	103616_at	AF100956	Col11a2	procollagen, type XI, alpha 2
41	100476_at	L20232	Ibsp	integrin binding sialoprotein
42	134747_at	AI662497	8430438D04Rik	RIKEN cDNA 8430438D04 gene
43	131792_s_at	AI850822		ESTs, Weakly similar to ENP1_MOUSE Ectonucleoside triphosphate diphosphohydrolase 1 (NTPDase1) (Ecto-ATP diphosphohydrolase) (ATPDase) (Lymphoid cell activation antigen) (CD39 antigen) (Ecto-apyrase) [M.musulus]
44	115962_at	AI615325		ESTs
45	102261_f_at	U30292	Col13a1	procollagen, type XIII, alpha 1
46	113640_at	AW045201	1110004G24Rik	RIKEN cDNA 1110004G24 gene
47	132364_i_at	AI594430		ESTs
48	132365_r_at	AI594430		ESTs
49	135131_at	AI508877		ESTs, Weakly similar to ITA2_MOUSE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GPIa) (Collagen receptor) (VLA-2 alpha chain) (CD49b) [M.musulus]
50	100484_at	X66473	Mmp13	matrix metalloproteinase 13
51	100336_s_at	L24431	Bglap2	bone gamma-carboxyglutamate protein 2
52	99957_at	X72795	Mmp9	matrix metalloproteinase 9
53	99903_at	AJ242625	Dmp1	dentin matrix protein 1
54	98482_at	X78936	Pthr	parathyroid hormone receptor
55	92796_at	J02980	Akp2	alkaline phosphatase 2, liver
56	93369_at	AB007848	Omd	Osteomodulin
57	93136_at	D78274	Dspg3	dermatan sulphate proteoglycan 3
58	**139531_at**	AW120795	**AW048930**	**expressed sequence AW048930**
59	97109_at	L00653	Mcpt7	mast cell protease 7
60	**117277_at**	AI850694	**MGC36197**	**hypothetical protein MGC36197**
61	**94930_at**	L07803	**Thbs2**	**thrombospondin 2**
62	104477_at	AW047643		ESTs
63	**92705_at**	U15566	**Tbx2**	**T-box 2**
64	**112062_at**	AA989939	**Colec12**	**collectin sub-family member 12**

BOLD genes were not found in cluster 3.

Appendix D

Table 1 | Genes Upregulated in Empty Found in the Statistical Analysis

Probe Set	Title	Gene Symbol
L29376_at		
Z11793_at	selenoprotein P, plasma, 1	SEPP1
HG1699-HT1704_s_at		
U06863_at	follistatin-like 1	FSTL1
Z14982_rna1_at		
HG4322-HT4592_at		
D14664_at	KIAA0022 gene product	KIAA0022
U13913_s_at	potassium large conductance calcium-activated channel, subfamily M, alpha member 1	KCNMA1
U40490_at	nicotinamide nucleotide transhydrogenase	NNT
X97324_at	adipose differentiation-related protein	ADFP
U30930_at	UDP glycosyltransferase 8 (UDP-galactose ceramide galactosyltransferase)	UGT8
U56402_s_at	suppressor of Ty 5 homolog (S. cerevisiae)	SUPT5H
X98176_at	caspase 8, apoptosis-related cysteine protease	CASP8
U03858_at	fms-related tyrosine kinase 3 ligand	FLT3LG
L13286_at	cytochrome P450, subfamily XXIV (vitamin D 24-hydroxylase)	CYP24
D13638_s_at		
HG3415-HT3598_at		
X95677_at		
M57710_at	lectin, galactoside-binding, soluble, 3 (galectin 3)	LGALS3
U32849_at	N-myc (and STAT) interactor	NMI
U18288_at	MHC class II transactivator	MHC2TA
U20758_rna1_at		
U08191_s_at	nuclear factor related to kappa B binding protein	NFRKB
M27968_s_at	fibroblast growth factor 2 (basic)	FGF2
X69636_at		
HG2999-HT4756_s_at		
AFFX-HUMTFRR/M11507_M_at	transferrin receptor (p90, CD71)	TFRC
U15655_at	Ets2 repressor factor	ERF
L07615_at	neuropeptide Y receptor Y1	NPY1R
D86976_at	minor histocompatibility antigen HA-1	HA-1
X91911_s_at	glioma pathogenesis-related protein	RTVP1
M59499_at		
Y07759_at	myosin VA (heavy polypeptide 12, myoxin)	MYO5A
U20240_at	CCAAT/enhancer binding protein (C/EBP), gamma	CEBPG
X72889_at	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2	SMARCA2
D84239_at	Fc fragment of IgG binding protein	FCGBP
M17733_at	thymosin, beta 4, X chromosome	TMSB4X
U03688_at	cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1 (glaucoma 3, primary infantile)	CYP1B1
X90824_s_at	upstream transcription factor 2, c-fos interacting	USF2
L07077_at	enoyl-Coenzyme A, hydratase/3-hydroxyacyl Coenzyme A dehydrogenase	EHHADH
U12424_s_at	glycerol-3-phosphate dehydrogenase 2 (mitochondrial)	GPD2
U09937_rna1_s_at		
HG2562-HT2658_s_at		

L08488_at	inositol polyphosphate-1-phosphatase	INPP1
D86983_at	Melanoma associated gene	D2S448
L14269_at	solute carrier family 18 (vesicular monoamine), member 2	SLC18A2
L08835_rna2_s_at		
D83703_at	peroxisomal biogenesis factor 6	PEX6
L75847_at	zinc finger protein 45 (a Kruppel-associated box (KRAB) domain polypeptide)	ZNF45
L34357_at	GATA binding protein 4	GATA4
U22961_s_at		
U66661_at	gamma-aminobutyric acid (GABA) A receptor, epsilon	GABRE
Z29083_at	trophoblast glycoprotein	TPBG
X74837_at	mannosidase, alpha, class 1A, member 1	MAN1A1
U21051_rna1_at		
L20591_at		
U16307_at	glioma pathogenesis-related protein	RTVP1
U31201_cds2_s_at		
U49742_at		
Z80781_at		
U69961_at	paired-like homeodomain transcription factor 2	PITX2
M87313_s_at		
D29810_at	endothelial and smooth muscle cell-derived neuropilin-like protein	ESDN
HG162-HT3165_at		
HG4126-HT4396_at		
AB000410_s_at	8-oxoguanine DNA glycosylase	OGG1
D50683_at	transforming growth factor, beta receptor II (70-80kD)	TGFBR2
U83115_at	absent in melanoma 1	AIM1
X84373_at	nuclear receptor interacting protein 1	NRIP1
X96584_at	nephroblastoma overexpressed gene	NOV
U40992_at	DnaJ (Hsp40) homolog, subfamily B, member 4	DNAJB4
L16895_at		
U16306_at	chondroitin sulfate proteoglycan 2 (versican)	CSPG2
X86428_s_at		
M76125_s_at	AXL receptor tyrosine kinase	AXL
M35878_at		
S75881_s_at	v-myb myeloblastosis viral oncogene homolog (avian)-like 1	MYBL1
U65533_s_at	regulator of nonsense transcripts 1	RENT1
X05908_at	annexin A1	ANXA1
L13329_at		
M34057_at	latent transforming growth factor beta binding protein 1	LTBP1
M24899_at	thyroid hormone receptor, alpha (erythroblastic leukemia viral (v-erb-a) oncogene homolog, avian)	THRA
U13021_s_at	caspase 2, apoptosis-related cysteine protease (neural precursor cell expressed, developmentally down-regulated 2)	CASP2
U84540_at		
U52513_at	interferon-induced protein with tetratricopeptide repeats 4	IFIT4
X82209_at	meningioma (disrupted in balanced translocation) 1	MN1
M60047_at	heparin-binding growth factor binding protein	HBP17
HG2007-HT2056_s_at		
M26901_s_at		
L19267_at	dystrophia myotonica-containing WD repeat motif	DMWD
X59656_at	v-crk sarcoma virus CT10 oncogene homolog (avian)-like	CRKL
AB002365_at	KIAA0367 protein	KIAA0367
X74039_at	plasminogen activator, urokinase receptor	PLAUR

L41351_at	protease, serine, 8 (prostasin)	PRSS8
D13631_s_at	Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6	ARHGEF6
M15059_at	Fc fragment of IgE, low affinity II, receptor for (CD23A)	FCER2
M77349_at	transforming growth factor, beta-induced, 68kD	TGFBI
X99268_at	twist homolog (acrocephalosyndactyly 3; Saethre-Chotzen syndrome) (Drosophila)	TWIST
Z26653_at	laminin, alpha 2 (merosin, congenital muscular dystrophy)	LAMA2
J03764_at		
X04729_s_at		
M77140_at	galanin	GAL
U33822_at	MAD1 mitotic arrest deficient-like 1 (yeast)	MAD1L1
X52221_at		
D32002_s_at	nuclear cap binding protein subunit 1, 80kD	NCBP1
M62402_at	insulin-like growth factor binding protein 6	IGFBP6
U62961_at	3-oxoacid CoA transferase	OXCT
AD000684_cds1_at		
M22324_at	alanyl (membrane) aminopeptidase (aminopeptidase N, aminopeptidase M, microsomal aminopeptidase, CD13, p150)	ANPEP
U70981_at	interleukin 13 receptor, alpha 2	IL13RA2

Table 2 | Genes Downregulated in Empty Found in the Statistical Analysis

Probe Set	Title	Gene Symbol
D31764_at	sorting nexin 17	SNX17
X52541_at	early growth response 1	EGR1
U72648_s_at		
K03021_at		
X55037_s_at	GATA binding protein 3	GATA3
V01512_rna1_at		
S69790_at		
AD000092_cds2_at		
X02761_s_at	fibronectin 1	FN1
HG3044-HT3742_s_at		
Y10615_at		
D63482_at	G protein-coupled receptor kinase-interactor 2	GIT2
U77968_at	neuronal PAS domain protein 1	NPAS1
L20971_at	phosphodiesterase 4B, cAMP-specific (phosphodiesterase E4 dunce homolog, Drosophila)	PDE4B
X84746_at		
L10665_at	guanine nucleotide binding protein (G protein), alpha activating activity polypeptide, olfactory type	GNAL
X06562_at	growth hormone receptor	GHR
X77307_at	5-hydroxytryptamine (serotonin) receptor 2B	HTR2B
U78180_at	amiloride-sensitive cation channel 2, neuronal	ACCN2
D10495_at	protein kinase C, delta	PRKCD
HG2663-HT2759_at		
D49817_at	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	PFKFB3
L11702_at	glycosylphosphatidylinositol specific phospholipase D1	GPLD1
Y00787_s_at	interleukin 8	IL8
U60062_at	fasciculation and elongation protein zeta 1 (zygin I)	FEZ1
X52009_s_at	glycine receptor, alpha 1 (startle disease/hyperekplexia, stiff man syndrome)	GLRA1
X69433_at	isocitrate dehydrogenase 2 (NADP+), mitochondrial	IDH2
L42450_at	pyruvate dehydrogenase kinase, isoenzyme 1	PDK1

HG2147-HT2217_r_at		
S77154_s_at	nuclear receptor subfamily 4, group A, member 2	NR4A2
L19314_at		
X75918_at	nuclear receptor subfamily 4, group A, member 2	NR4A2
D79990_at	Ras association (RalGDS/AF-6) domain family 2	RASSF2
M69225_at		
Z17240_at	high-mobility group (nonhistone chromosomal) protein 2	HMG2
L03840_s_at	fibroblast growth factor receptor 4	FGFR4
HG415-HT415_at		
D14134_at	RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae)	RAD51
U09646_at		
X02530_at	small inducible cytokine subfamily B (Cys-X-Cys), member 10	SCYB10
L17075_s_at	activin A receptor type II-like 1	ACVRL1
U90908_at	hypothetical protein from clones 23549 and 23762	LOC58504
HG3125-HT3301_s_at		
U19523_at	GTP cyclohydrolase 1 (dopa-responsive dystonia)	GCH1
D82343_at	olfactomedin 1	OLFM1
U61741_at		
X93498_at	SH3 domain binding glutamic acid-rich protein	SH3BGR
U73936_at	jagged 1 (Alagille syndrome)	JAG1
M65291_at	interleukin 12A (natural killer cell stimulatory factor 1, cytotoxic lymphocyte maturation factor 1, p35)	IL12A
D87075_at	solute carrier family 23 (nucleobase transporters), member 1	SLC23A1
L13197_at		
U96114_at	Nedd-4-like ubiquitin-protein ligase	WWP2
L19871_at	activating transcription factor 3	ATF3
M22898_at		
L08096_s_at	tumor necrosis factor (ligand) superfamily, member 7	TNFSF7
X16832_at	cathepsin H	CTSH
U15177_at		
U09414_at	zinc finger protein 137 (clone pHZ-30)	ZNF137
D87074_at	KIAA0237 gene product	KIAA0237
U01317_cds6_at		
U12767_at	nuclear receptor subfamily 4, group A, member 3	NR4A3
HG3123-HT3299_at		
X80497_at	phosphorylase kinase, alpha 2 (liver)	PHKA2
U59914_at	MAD, mothers against decapentaplegic homolog 6 (Drosophila)	MADH6
X15306_rna1_at		
S82471_s_at		
U35459_at	serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 10	SERPINB10
U60115_at	four and a half LIM domains 1	FHL1
D21267_at	synaptosomal-associated protein, 25kD	SNAP25
U47926_at	hypothetical protein B	HSU47926
S81916_at		
U28833_at	Down syndrome critical region gene 1	DSCR1
HG3494-HT3688_at		
X82240_rna1_at	T-cell leukemia/lymphoma 1A	TCL1A
U62531_at	solute carrier family 4, anion exchanger, member 2 (erythrocyte membrane protein band 3-like 1)	SLC4A2
U89995_at	forkhead box E1 (thyroid transcription factor 2)	FOXE1
M27318_f_at	interferon, alpha 4	IFNA4

D85376_at		
X15414_at	aldo-keto reductase family 1, member B1 (aldose reductase)	AKR1B1
M14123_xpt2_at		
J02854_at	myosin, light polypeptide 9, regulatory	MYL9
HG2981-HT3127_s_at		
Z29067_at	NIMA (never in mitosis gene a)-related kinase 3	NEK3
M30607_s_at	zinc finger protein, Y-linked	ZFY
U28055_at	macrophage stimulating, pseudogene 9	MSTP9
L39211_at	carnitine palmitoyltransferase I, liver	CPT1A
D49372_s_at	small inducible cytokine subfamily A (Cys-Cys), member 11 (eotaxin)	SCYA11
M55418_at		

Table 3 | Genes of Cluster G2. High in Empty, Low in Mutants.

Name	Description
D14664_at	D14664, class A, 20 probes, 20 in D14664 3255-3639, Human mRNA for KIAA0022 gene, complete cds
D31886_at	D31886, class A, 20 probes, 20 in D31886 3076-3592, Human mRNA for KIAA0066 gene, partial cds
D49387_at	D49387, class A, 20 probes, 20 in D49387 401-917, Human mRNA for NADP dependent leukotriene b4 12-hydroxydehydrogenase, partial cds. /gb=D49387 /ntype=RNA
D50683_at	D50683, class A, 20 probes, 20 in D50683 5296-5680, Human mRNA for TGF-betaIIR alpha, complete cds
D78611_at	D78611, class A, 20 probes, 20 in D78611 1893-2331, Human MEST mRNA, complete cds
D83018_at	D83018, class A, 20 probes, 20 in D83018 2645-3149, Human mRNA for nel-related protein 2, complete cds
D86640_at	D86640, class A, 20 probes, 20 in D86640 2374-2902, Human mRNA for stac, complete cds
D86976_at	D86976, class A, 20 probes, 20 in D86976 3592-4060, Human mRNA for KIAA0223 gene, partial cds
D86983_at	D86983, class A, 20 probes, 20 in D86983 5131-5485, Human mRNA for KIAA0230 gene, partial cds
HG3415-HT3598_at	Poliovirus Receptor
HG4126-HT4396_at	Zinc Finger Protein Hzf4
HG4321-HT4591_at	Ahnak-Related Sequence
J03258_at	J03258, class A, 20 probes, 20 in J03258mRNA 4003-4561, Human vitamin D receptor mRNA, complete cds
J04970_at	J04970, class A, 20 probes, 20 in J04970 1397-1715, Human carboxypeptidase M, 3 end
L13720_at	L13720, class A, 20 probes, 20 in L13720 1860-2436, Homo sapiens growth-arrest-specific protein (gas) mRNA, complete cds
L19872_at	L19872, class A, 20 probes, 20 in L19872 4756-5059, Human AH-receptor mRNA, complete cds
L23116_at	L23116, class A, 20 probes, 20 in L23116 3296-3644, Homo sapiens galactocerebrosidase (GALC) mRNA, complete cds
L40377_at	L40377, class A, 20 probes, 20 in L40377mRNA 766-1276, Homo sapiens cytoplasmic antiproteinase 2 (CAP2) mRNA, complete cds
L41351_at	L41351, class A, 20 probes, 20 in L41351mRNA 1269-1695, Homo sapiens prostasin mRNA, complete cds
L75847_at	L75847, class A, 20 probes, 20 in L75847 1808-2330, Human zinc finger protein 45 (ZNF45) mRNA, complete cds
M13194_at	M13194, class A, 20 probes, 20 in M13194mRNA 586-1006, Human excision repair protein (ERCC1) mRNA, complete cds, clone pcDE
M23668_at	M23668, class A, 20 probes, 20 in M23668exon 743-1271, Homo sapiens adrenodoxin gene
M24899_at	M24899, class A, 20 probes, 20 in M24899 1750-2284, Human triiodothyronine (ear7) mRNA, complete cds
M37825_at	M37825, class A, 20 probes, 20 in M37825 624-1044, Human fibroblast growth factor-5 (FGF-5) mRNA, complete cds
M55420_at	M55420, class C, 20 probes, 18 in all_M55420 605-897: 2 in M55420cds 109-140, Human IgE chain, last 2 exons

M57710_at	M57710, class A, 20 probes, 20 in M57710 355-865, Human IgE-binding protein (epsilon-BP) mRNA, complete cds
M60047_at	M60047, class A, 20 probes, 20 in M60047 641-1097, Human heparin binding protein (HBp17) mRNA, complete cds
M96684_at	M96684, class A, 20 probes, 20 in M96684 609-867, H.sapiens Pur (pur-alpha) mRNA, complete cds
U02082_at	U02082, class A, 20 probes, 20 in U02082 1643-2201, Human guanine nucleotide regulatory protein (tim1) mRNA, complete cds
U03688_at	U03688, class A, 20 probes, 20 in U03688 4501-5047, Human dioxin-inducible cytochrome P450 (CYP1B1) mRNA, complete cds
U06863_at	U06863, class A, 20 probes, 20 in U06863 1416-1938, Human follistatin-related protein precursor mRNA, complete cds
U09770_at	U09770, class A, 20 probes, 20 in U09770 61-391, Human cysteine-rich heart protein (hCRHP) mRNA, complete cds
U19718_at	U19718, class A, 20 probes, 20 in U19718 479-947, Human microfibril-associated glycoprotein (MFAP2) mRNA, complete cds
U30998_at	U30998, class A, 20 probes, 20 in U30998 43-166, Human (nmd) mRNA, 3 UTR. /gb=U30998 /ntype=RNA
U39840_at	U39840, class A, 20 probes, 20 in U39840 2313-2823, Human hepatocyte nuclear factor-3 alpha (HNF-3 alpha) mRNA, complete cds
U40992_at	U40992, class A, 20 probes, 20 in U40992 839-1175, Human heat shock protein hsp40 homolog mRNA, complete cds
U46499_at	U46499, class C, 20 probes, 20 not in GB record, Human microsomal glutathione transferase (GST12) gene, 5 sequence
U63717_at	U63717, class A, 20 probes, 20 in U63717 402-852, Human osteoclast stimulating factor mRNA, complete cds
U70981_at	U70981, class A, 20 probes, 20 in U70981 749-1283, Human interleukin-13 receptor mRNA, complete cds
U78095_at	U78095, class A, 20 probes, 20 in U78095 942-1434, Human Placental bikunin mRNA, complete cds
U79263_at	U79263, class A, 20 probes, 20 in U79263 995-1535, Human clone 23760 mRNA, partial cds.
U83115_at	U83115, class A, 20 probes, 20 in U83115 6327-6753, Human non-lens beta gamma-crystallin like protein (AIM1) mRNA, partial cds
U89916_at	U89916, class A, 20 probes, 20 in U89916 375-879, Homo sapiens putative OSP like protein mRNA, partial cds.
X59656_at	X59656, class C, 20 probes, 20 in all_X59656 1286-1827, H.sapiens crk-like gene CRKL
X61615_at	X61615, class B, 20 probes, 13 in X61615cds 2830-3160: 7 in reverseSequence, 3482-3548, H.sapiens mRNA for leukemia inhibitory factor (LIF) receptor
X66922_at	X66922, class A, 20 probes, 18 in X66922cds 362-728: 2 in reverseSequence, 848-872, H.sapiens mRNA for myo-insositol monophosphatase
X69636_at	X69636, class C, 20 probes, 20 in all_X69636 1268-1951, H.sapiens mRNA sequence (15q11-13)
X70394_at	X70394, class C, 20 probes, 20 in all_X70394 2591-3156, H.sapiens OZF mRNA
X74837_at	X74837, class C, 20 probes, 20 in all_X74837 2811-3196, H.sapiens HUMM9 mRNA
X84373_at	X84373, class C, 20 probes, 20 in all_X84373 6655-7208, H.sapiens mRNA for nuclear factor RIP140
X87212_at	X87212, class C, 20 probes, 20 in all_X87212 1273-1772, H.sapiens mRNA for cathepsin C
X96584_at	X96584, class C, 20 probes, 20 in all_X96584 1444-1961, H.sapiens mRNA for NOV protein
X97324_at	X97324, class A, 20 probes, 20 in X97324cds 749-1277, H.sapiens mRNA for adipophilin. /gb=X97324 /ntype=RNA
Y07596_at	Y07596, class B, 20 probes, 11 in Y07596cds 1035-1149: 9 in reverseSequence, 1173-1509, H.sapiens mRNA for GPI8 protein
Y08136_at	Y08136, class B, 20 probes, 12 in Y08136cds 292-496: 8 in reverseSequence, 520-820, H.sapiens mRNA for ASM-like phosphodiesterase 3a
Y11251_at	Y11251, class A, 20 probes, 20 in Y11251 4297-4822, H.sapiens mRNA for novel member of serine-arginine domain protein, SRrp129
Y11681_at	Y11681, class C, 20 probes, 20 in all_Y11681 529-1040, Homo sapiens mRNA for mitochondrial ribosomal protein S12. /gb=Y11681 /ntype=RNA
L07919_at	L07919, class A, 20 probes, 20 in L07919 1386-1779, Human homeodomain protein DLX-2 mRNA, 3 end
U57627_at	U57627, class A, 20 probes, 20 in U57627 4598-5078, Human fetal brain oculocerebrorenal syndrome (OCRL1) mRNA, complete cds
X92368_at	X92368, class A, 20 probes, 20 in X92368mRNA 5695-6187, H.sapiens ncx1 gene (exon 1). /gb=X92368 /ntype=DNA /annot=mRNA
X95525_at	X95525, class C, 20 probes, 20 in all_X95525 2560-3071, H.sapiens mRNA for

	TAFII100 protein
D13631_s_at	D13631, class A, 20 probes, 20 in D13631 2795-3373, Human mRNA for KIAA0006 gene, complete cds
D13638_s_at	D13638, class A, 20 probes, 20 in D13638 5003-5557, Human mRNA for KIAA0013 gene, complete cds.
HG1699-HT1704_s_at	Epimorphin - Also Represents: D14582
D16611_s_at	D16611, class A, 20 probes, 20 in D16611 1726-2299, Human mRNA for coproporphyrinogen oxidase, complete cds
D86062_s_at	D86062, class A, 19 probes, 19 in D86062 286-862, Human mRNA for KNP-Ib, complete cds
U65533_s_at	U65533, class A, 20 probes, 20 in U65533 3076-3620, Human regulator of nonsense transcript stability (RENT1) mRNA, complete cds
M76125_s_at	M76125, class A, 20 probes, 20 in M76125 2612-3170, Human tyrosine kinase receptor (axl) mRNA, complete cds
HG2007-HT2056_s_at	Proto-Oncogene Sno, Alt. Splice N - Also Represents: HG2007-HT5112
L08835_rna2_s_at	L08835, class A, 20 probes, 20 in L08835mRNA#1 3166-3367, DM kinase gene (myotonic dystrophy kinase) extracted from Human myotonic dystrophy kinase (DM kinase) gene, complete cds
HG2562-HT2658_s_at	A-Myb (Gb:X13294) - Also Represents: X66087
HG3914-HT4184_s_at	Cell Division Cycle Protein 2-Related Protein Kinase (Pisslre) - Also Represents: X78342
L76528_s_at	L76528, class A, 20 probes, 20 in L76528exon 146-615, Homo sapiens presenilin 1 (PS1; S182) gene
X63131_s_at	X63131, class C, 20 probes, 20 in all_X63131 1996-2179, H.sapiens My1 (PML) mRNA
X74929_s_at	X74929, class C, 20 probes, 20 in all_X74929 1365-1706, H.sapiens KRT8 mRNA for keratin 8
U12424_s_at	U12424, class A, 20 probes, 20 in U12424 2016-2564, Human mitochondrial glycerol-3-phosphate dehydrogenase mRNA, complete cds
U13021_s_at	U13021, class A, 19 probes, 19 in U13021 844-1392, Human positive regulator of programmed cell death ICH-1L (Ich-1) mRNA, complete cds
U19252_s_at	U19252, class A, 20 probes, 20 in U19252 4495-5045, Human putative transmembrane protein mRNA, complete cds
U75309_s_at	U75309, class A, 20 probes, 20 in U75309 1813-2376, Human TBP-associated factor (hTAFII100) mRNA, partial cds
X83492_s_at	X83492, class C, 15 probes, 15 in all_X83492 418-500, H.sapiens mRNA for Fas/Apo-1 (clone pCRTM11-Fasdelta(4,7)). /gb=X83492 /ntype=RNA, H.sapiens mRNA for Fas/Apo-1 (clone pCRTM11-Fasdelta(4,7)). /gb=X83492 /ntype=RNA
X83490_s_at	X83490, class A, 20 probes, 19 in X83490exon 3-34: 1 in reverseSequence, 389, H.sapiens mRNA for Fas/Apo-1 (clone pCRTM11-Fasdelta(3,4)). /gb=X83490 /ntype=RNA
X86428_s_at	X86428, class B, 20 probes, 14 in X86428cds 626-920: 6 not in GB record, H.sapiens gene for phosphotyrosyl phosphatase activator (exon 1)
HG3635-HT3845_f_at	Zinc Finger Protein, Kruppel-Like
HG4322-HT4592_at	Tubulin, Beta
K02405_f_at	K02405, class C, 20 probes, 19 in all_K02405 5550-7761: 1 in K02405cds 778, Human MHC class II HLA-DC-3-beta gene (DR3,3)
M71243_f_at	M71243, class B, 16 probes, 14 in M71243mRNA 25-38: 2 not in GB record, Human glycophorin Sta (type A) exons 3 and 4, partial. /gb=M71243 /ntype=DNA /annot=exon
U18934_at	U18934, class A, 20 probes, 20 in U18934 4229-4311, Human receptor tyrosine kinase (DTK) mRNA, complete cds
Z80781_at	Z80781, class C, 20 probes, 20 in all_Z80781 583-748, H.sapiens H2B/j gene.
X01703_at	X01703, class A, 20 probes, 20 in X01703exon#4 929-1151, Human gene for alpha-tubulin (b alpha 1)
U43944_at	U43944, class A, 20 probes, 20 in U43944 1705-1978, Human breast cancer cytosolic NADP(+)-dependent malic enzyme mRNA, partial cds
X95632_s_at	X95632, class C, 20 probes, 20 in all_X95632 1680-1784, H.sapiens mRNA for Arg protein tyrosine kinase-binding protein
Y09392_s_at	Y09392, class A, 20 probes, 20 in Y09392exon#4 364-884, H.sapiens mRNA for WSL-LR, WSL-S1 and WSL-S2 proteins
U72263_s_at	U72263, class A, 20 probes, 20 in U72263 2410-2931, Human multiple exostoses type II protein EXT2.I mRNA, complete cds. /gb=U72263 /ntype=RNA
U56402_s_at	U56402, class A, 20 probes, 20 in U56402 2969-3471, Human chromatin structural protein homolog (SUPT5H) mRNA, complete cds
U31201_cds2_s_at	U31201, class A, 20 probes, 20 in U31201mRNA 4592-5106, Human laminin gamma2 chain gene (LAMC2), Human laminin gamma2 chain gene (LAMC2)

X91911_s_at	X91911, class A, 20 probes, 18 in X91911cds 321-711: 2 in reverseSequence, 912-950, H.sapiens mRNA for RTVP-1 protein
U08191_s_at	U08191, class A, 20 probes, 20 in U08191 4687-5220, Human R kappa B mRNA, complete cds
S75881_s_at	S75881, class A, 20 probes, 20 in S75881 234-719, A-myb=DNA-binding transactivator {3 region} [human, CCRF-CEM T-leukemia line, mRNA Partial, 831 nt]
X90824_s_at	X90824, class C, 19 probes, 19 in all_X90824 828-1337, H.sapiens mRNA for USF2a & USF2b, clone P9DH
M20778_s_at	M20778, class A, 20 probes, 20 in M20778 401-974, Homo sapien, alpha-3 (VI) collagen
M26901_s_at	M26901, class A, 20 probes, 16 in M26901cds 808-1187: 4 in reverseSequence, 218-293, Human renin gene
M27968_s_at	M27968, class A, 20 probes, 20 in M27968mRNA 3289-3658, Human basic fibroblast growth factor (FGF) mRNA, complete cds
X04729_s_at	X04729, class C, 19 probes, 19 in all_X04729 2-263, Human mRNA for plasminogen activator inhibitor type 1 N-terminus. /gb=X04729 /ntype=RNA
U09937_rna1_s_at	U09937, class A, 20 probes, 20 in U09937mRNA 1176-1581, urokinase-type plasminogen activator receptor gene extracted from Human urokinase-type plasminogen receptor
D32002_s_at	D32002, class A, 20 probes, 20 in D32002 2454-3001, Human mRNA for nuclear cap binding protein, complete cds
D26155_s_at	D26155, class A, 20 probes, 20 in D26155 4647-5214, Human mRNA for transcriptional activator hSNF2a, complete cds
X82209_at	X82209, class A, 20 probes, 20 in X82209 7019-7511, H.sapiens MN1 mRNA
X72889_at	X72889, class C, 20 probes, 20 in all_X72889 5441-5844, H.sapiens hbrm mRNA
X66087_at	X66087, class C, 20 probes, 20 in all_X66087 3046-3563, H.sapiens a-myb mRNA
M15169_at	M15169, class A, 20 probes, 16 in M15169mRNA#1 1704-1950: 4 in reverseSequence, 3390-3408, Human beta-2-adrenergic receptor mRNA, complete cds
HG162-HT3165_at	Tyrosine Kinase, Receptor Axl, Alt. Splice 2
D29805_at	D29805, class A, 20 probes, 20 in D29805 3485-3995, Human mRNA for beta-1,4-galactosyltransferase, complete cds
Z29083_at	Z29083, class C, 20 probes, 20 in all_Z29083 1644-2023, H.sapiens 5T4 gene for 5T4 Oncofetal antigen
Z26653_at	Z26653, class A, 20 probes, 16 in Z26653cds 8896-9286: 4 in reverseSequence, 9383-9509, H.sapiens mRNA for laminin M chain (merosin)
Z14982_rna1_at	Z14982, class A, 20 probes, 20 in Z14982mRNA#1 616-1150, MHC-encoded proteasome subunit gene LAMP7-E1 gene (proteasome subunit LMP7) extracted from H.sapiens gene for major histocompatibility complex encoded proteasome subunit LMP7
Y09321_at	Y09321, class A, 20 probes, 17 in Y09321cds 1961-2375: 3 in reverseSequence, 2423-2501, H.sapiens TAFII105 mRNA, partial
Y07759_at	Y07759, class C, 20 probes, 20 in all_Y07759 5956-6377, H.sapiens mRNA for myosin heavy chain 12
X99268_at	X99268, class C, 20 probes, 20 in all_X99268 928-1367, H.sapiens mRNA for B-HLH DNA binding protein
X98172_at	X98172, class C, 20 probes, 20 in all_X98172 2240-2754, H.sapiens mRNA for MACH-alpha-1 protein
X89059_at	X89059, class C, 20 probes, 20 in all_X89059 722-1203, H.sapiens mRNA for unknown protein expressed in macrophages
X74039_at	X74039, class C, 20 probes, 20 in all_X74039 805-1058, H.sapiens mRNA for urokinase plasminogen activator receptor
X52221_at	X52221, class A, 20 probes, 20 in X52221mRNA 1674-2244, H.sapiens ERCC2 gene, exons 1 & 2 (partial)
X05908_at	X05908, class B, 20 probes, 11 in X05908cds 814-1012: 9 in reverseSequence, 1110-1338, Human mRNA for lipocortin
X05196_at	X05196, class A, 20 probes, 18 in X05196exon#9 2-458: 1 in reverseSequence, 3199: 1 not in GB record, Human aldolase C gene
U84573_at	U84573, class A, 20 probes, 20 in U84573 2882-3422, Homo sapiens lysyl hydroxylase isoform 2 (PLOD2) mRNA, complete cds
U69961_at	U69961, class A, 20 probes, 20 in U69961 1565-1997, Human solurshin (RGS) mRNA, complete cds
U66661_at	U66661, class A, 20 probes, 20 in U66661 2656-3082, Human GABA-A receptor epsilon subunit mRNA, complete cds
U64871_at	U64871, class A, 20 probes, 17 in U64871cds 870-1212: 3 in reverseSequence, 1665-1773, Human putative G protein-coupled receptor (GPR19) gene, complete cds
U63289_at	U63289, class A, 20 probes, 20 in U63289 1548-2010, Human RNA-binding protein CUG-BP/hNab50 (NAB50) mRNA, complete cds.
U62961_at	U62961, class A, 20 probes, 20 in U62961 2749-3241, Human succinyl CoA:3-oxoacid CoA transferase precursor (OXCT) mRNA, complete cds

U52513_at	U52513, class A, 20 probes, 20 in U52513 1318-1642, Human RIG-G mRNA, complete cds
U40490_at	U40490, class A, 20 probes, 20 in U40490 3673-4177, Human nicotinamide nucleotide transhydrogenase mRNA, nuclear gene encoding mitochondrial protein, complete cds
U35048_at	U35048, class A, 20 probes, 20 in U35048 1159-1675, Human TSC-22 protein mRNA, complete cds
U33822_at	U33822, class A, 20 probes, 20 in U33822 2053-2563, Human tax1-binding protein TXBP181 mRNA, complete cds
U32849_at	U32849, class A, 20 probes, 20 in U32849 867-1383, Human Hou mRNA, complete cds
U32645_at	U32645, class A, 20 probes, 20 in U32645 3566-4112, Human myeloid elf-1 like factor (MEF) mRNA, complete cds
U32315_at	U32315, class A, 20 probes, 20 in U32315 1374-1842, Human syntaxin 3 mRNA, complete cds
U30930_at	U30930, class A, 20 probes, 20 in U30930 1877-2423, Human UDP-Galactose ceramide galactosyl transferase (CGT) mRNA, complete cds
U28811_at	U28811, class A, 20 probes, 20 in U28811 3404-3866, Human cysteine-rich fibroblast growth factor receptor (CFR-1) mRNA, complete cds
U27655_at	U27655, class A, 20 probes, 20 in U27655 2169-2577, Human RGP3 mRNA, complete cds
U20240_at	U20240, class A, 20 probes, 20 in U20240 448-898, Human C/EBP gamma mRNA, complete cds
U16306_at	U16306, class A, 20 probes, 20 in U16306 10722-11142, Human chondroitin sulfate proteoglycan versican V0 splice-variant precursor peptide mRNA, complete cds
U15655_at	U15655, class A, 20 probes, 20 in U15655 2102-2576, Human ets domain protein ERF mRNA, complete cds
U14383_at	U14383, class A, 20 probes, 20 in U14383 958-1372, Human mucin (MUC8) mRNA, partial cds
U01062_at	U01062, class A, 20 probes, 20 in U01062mRNA 8334-8778, Human type 3 inositol 1,4,5-trisphosphate receptor (ITPR3) mRNA, complete cds
S67970_at	S67970, class A, 20 probes, 20 in S67970 962-1538, ZNF75=KRAB zinc finger [human, lung fibroblast, mRNA, 1563 nt]
M79462_at	M79462, class A, 20 probes, 20 in M79462 3853-4333, Human PML-1 mRNA, complete CDS
M77349_at	M77349, class A, 20 probes, 20 in M77349 2102-2642, Human transforming growth factor-beta induced gene product (BIGH3) mRNA, complete cds
M77140_at	M77140, class A, 20 probes, 20 in M77140 91-409, H.sapiens pro-galanin mRNA, 3 end
M62402_at	M62402, class A, 20 probes, 20 in M62402 453-927, Human insulin-like growth factor binding protein 6 (IGFBP6) mRNA, complete cds
M59911_at	M59911, class A, 20 probes, 20 in M59911 4048-4612, Human integrin alpha-3 chain mRNA, complete cds
M35878_at	M35878, class A, 20 probes, 20 in M35878exon#4 1993-2443, Human insulin-like growth factor-binding protein-3 gene, complete cds, clone HL1006d
M35416_at	M35416, class A, 20 probes, 20 in M35416mRNA 864-1302, Human GTP-binding protein (RALB) mRNA, complete cds
M34057_at	M34057, class A, 20 probes, 20 in M34057 4720-5044, Human transforming growth factor-beta 1 binding protein mRNA, complete cds
M22324_at	M22324, class A, 20 probes, 20 in M22324 2954-3416, Human aminopeptidase N/CD13 mRNA encoding aminopeptidase N, complete cds
M17733_at	M17733, class A, 20 probes, 20 in M17733mRNA 13-505, Human thymosin beta-4 mRNA, complete cds
M15059_at	M15059, class A, 20 probes, 20 in M15059mRNA 1025-1487, Human Fc-epsilon receptor (IgE receptor) mRNA, complete cds (H107 epitope)
M13755_at	M13755, class A, 20 probes, 20 in M13755mRNA 33-591, Human interferon-induced 17-kDa/15-kDa protein mRNA, complete cds
L40027_at	L40027, class A, 20 probes, 20 in L40027mRNA 1586-2132, Homo sapiens glycogen synthase kinase 3 mRNA, complete cds
L35251_rna1_at	L35251, class A, 20 probes, 20 in L35251mRNA 801-1281, Homo sapiens extracellular matrix protein (MFAP3) gene, complete cds
L19267_at	L19267, class A, 20 probes, 20 in L19267 2335-2755, Homo sapiens 59 protein mRNA, 3 end
L13329_at	L13329, class A, 20 probes, 20 in L13329exon 434-938, Homo sapiens iduronate-2-sulfatase (IDS) gene
L10405_at	L10405, class A, 20 probes, 20 in L10405 1364-1910, Homo sapiens DNA binding protein for surfactant protein B mRNA, complete cds. /gb=L10405 /ntype=RNA
L08488_at	L08488, class A, 20 probes, 20 in L08488 1206-1644, Human inositol polyphosphate 1-phosphatase mRNA, complete cds
K01396_at	K01396, class A, 20 probes, 20 in K01396mRNA 769-1201, Human alpha-1-

	antitrypsin mRNA, complete cds
J04543_at	J04543, class A, 20 probes, 20 in J04543 1215-1725, Human synexin mRNA, complete cds
J03764_at	J03764, class C, 20 probes, 20 in all_J03764 14604-15049, Human, plasminogen activator inhibitor-1 gene, exons 2 to 9
J03589_at	J03589, class C, 20 probes, 20 in all_J03589 2962-3443, Human ubiquitin-like protein (GdX) gene, complete cds
HG511-HT511_at	Ras Inhibitor Inf
HG4194-HT4464_at	Sodium/Hydrogen Exchanger 5
HG172-HT3924_at	Spermidine/Spermine N1-Acetyltransferase, Alt. Splice 2
D84239_at	D84239, class A, 20 probes, 20 in D84239 15949-16339, Human mRNA for IgG Fc binding protein, complete cds
D83703_at	D83703, class A, 20 probes, 20 in D83703 2605-3169, Human mRNA for peroxisome assembly factor-2, complete cds
D63391_at	D63391, class A, 20 probes, 20 in D63391 341-773, Human mRNA for platelet activating factor acetylhydrolase IB gamma-subunit, complete cds
D50840_at	D50840, class A, 20 probes, 20 in D50840 1048-1474, Human mRNA for ceramide glucosyltransferase, complete cds
D29810_at	D29810, class A, 20 probes, 20 in D29810 835-1363, Human mRNA for unknown product, partial cds
D25217_at	D25217, class A, 20 probes, 20 in D25217 2864-3410, Human mRNA for KIAA0027 gene, partial cds
AD000684_cds1_at	AD000684, class B, 20 probes, 11 in AD000684cds#1 934-1252: 9 in reverseSequence, 16809-17037, LISCH7 gene (liver-specific bHLH-Zip transcription factor) extracted from Homo sapiens DNA from chromosome 19-cosmid R30879 containing USF2, genomic sequence
AC002115_rna2_at	AC002115, class A, 20 probes, 20 in AC002115mRNA#1 932-1448, COX6B gene (COXG) extracted from Human DNA from overlapping chromosome 19 cosmids R31396, F25451, and R31076 containing COX6B and UPKA, genomic sequence, COX6B gene (COXG) extracted from Human D
AB002365_at	AB002365, class A, 20 probes, 20 in AB002365 5053-5617, Human mRNA for KIAA0367 gene, partial cds. /gb=AB002365 /ntype=RNA

Table 4 | Genes of Cluster G3. Low in Empty, High in Mutants.

Name	Description
D14874_at	D14874, class A, 20 probes, 20 in D14874 908-1406, Human mRNA for adrenomedullin, complete cds
D32050_at	D32050, class A, 20 probes, 20 in D32050 2761-3307, Human mRNA for alanyl-tRNA synthetase, complete cds
D38491_at	D38491, class A, 20 probes, 20 in D38491 298-808, Human mRNA for KIAA0117 gene, partial cds
D49817_at	D49817, class A, 20 probes, 20 in D49817 1233-1725, Human mRNA for fructose 6-phosphate,2-kinase/fructose 2,6-bisphosphatase, complete cds
D79983_at	D79983, class A, 20 probes, 20 in D79983 5024-5498, Human mRNA for KIAA0161 gene, complete cds
D79990_at	D79990, class A, 20 probes, 20 in D79990 5065-5383, Human mRNA for KIAA0168 gene, complete cds
D79994_at	D79994, class A, 20 probes, 20 in D79994 4227-4749, Human mRNA for KIAA0172 gene, partial cds
D87074_at	D87074, class A, 20 probes, 20 in D87074 6650-7184, Human mRNA for KIAA0237 gene, complete cds
D87075_at	D87075, class A, 20 probes, 20 in D87075 5013-5469, Human mRNA for KIAA0238 gene, partial cds
HG174-HT174_at	Desmoplakin I
HG2810-HT2921_at	Homeotic Protein Pl2
HG3494-HT3688_at	Nuclear Factor Nf-Il6
HG415-HT415_at	Lectin, Galactoside-Binding, Soluble, 2
J03909_at	J03909, class A, 20 probes, 20 in J03909 461-995, Human gamma-interferon-inducible protein (IP-30) mRNA, complete cds
K03021_at	K03021, class A, 20 probes, 20 in K03021exon#14 343-853, Human tissue plasminogen activator (PLAT) gene, complete cds
M64936_at	M64936, class A, 20 probes, 20 in M64936 2808-3264, Homo sapiens retinoic acid-inducible endogenous retroviral DNA

M77836_at	M77836, class A, 20 probes, 20 in M77836 1239-1749, Human pyrroline 5-carboxylate reductase mRNA, complete cds
S77576_at	S77576, class A, 20 probes, 20 in S77576 3-60, ERV9 reverse transcriptase homolog {clone RT18} [human, multiple sclerosis, brain plaques, mRNA Partial, 84 nt]. /gb=S77576 /ntype=RNA
U14910_at	U14910, class A, 20 probes, 20 in U14910 910-1360, Human RPE-retinal G protein-coupled receptor (rgr) mRNA, complete cds
U15177_at	U15177, class C, 20 probes, 20 in all_U15177 2291-2724, Human cosmid CRI-JC2015 at D10S289 in 10sp13
U28833_at	U28833, class A, 20 probes, 20 in U28833 1571-2075, Human Down syndrome critical region protein (DSCR1) mRNA, complete cds
U31382_at	U31382, class A, 20 probes, 20 in U31382 69-621, Human G protein gamma-4 subunit mRNA, complete cds
U59111_at	U59111, class A, 20 probes, 20 in U59111 892-1444, Human dermatan sulfate proteoglycan 3 (DSPG3) mRNA, complete cds
U62531_at	U62531, class A, 20 probes, 20 in U62531 3465-4029, Human AE2 anion exchanger (SLC4A2) mRNA, complete cds
U66702_at	U66702, class A, 20 probes, 20 in U66702 4190-4616, Human phogrin mRNA, complete cds.
U75308_at	U75308, class A, 20 probes, 20 in U75308 3654-4092, Human TBP-associated factor (hTAFII130) mRNA, partial cds
U78180_at	U78180, class A, 20 probes, 20 in U78180 3340-3880, Human sodium channel 2 (hBNaC2) mRNA, alternatively spliced, complete cds
X12453_at	X12453, class A, 20 probes, 20 in X12453mRNA 993-1539, Human mRNA for retinal S-antigen (48 KDa protein)
X15306_rna1_at	X15306, class A, 20 probes, 20 in X15306mRNA 3269-3707, H.sapiens NF-H gene, exon 1 (and joined CDS)
X15414_at	X15414, class C, 20 probes, 20 in all_X15414 844-1349, Human mRNA for aldose reductase (EC 1.1.1.2)
X56667_at	X56667, class A, 20 probes, 20 in X56667mRNA 915-1341, Human mRNA for calretinin
X77307_at	X77307, class B, 20 probes, 13 in X77307cds 1244-1382: 7 in reverseSequence, 1491-1701, H.sapiens mRNA for 5-HT2B serotonin receptor
X84746_at	X84746, class A, 20 probes, 20 in X84746cds 544-1012, H.sapiens Histo-blood group AB0 gene, exon 1
X91504_at	X91504, class C, 20 probes, 20 in all_X91504 970-1523, H.sapiens mRNA for ARP1 protein
L00058_at	L00058, class C, 20 probes, 20 in all_L00058 470-855, Human (GH) germline c-myc proto-oncogene, 5 flank
L11702_at	L11702, class A, 20 probes, 20 in L11702 2837-3335, Human phospholipase D mRNA, complete cds
S76638_at	S76638, class A, 20 probes, 20 in S76638 2553-3003, p50-NF-kappa B homolog [human, peripheral blood T cells, mRNA, 3113 nt]
L27624_s_at	L27624, class A, 20 probes, 20 in L27624 373-917, Homo sapiens tissue factor pathway inhibitor-2 mRNA, complete cds
HG3044-HT3742_s_at	Fibronectin, Alt. Splice 1 - Also Represents: HG3044-HT2527, X02761
X02761_s_at	X02761, class C, 20 probes, 20 in all_X02761 7082-7646, Human mRNA for fibronectin (FN precursor)
HG3085-HT3254_s_at	Phosphodiesterase - Also Represents: U02882
HG3125-HT3301_s_at	Estrogen Receptor (Gb:S67777) - Also Represents: X03635
L25878_s_at	L25878, class A, 20 probes, 20 in L25878 1092-1657, Homo sapiens p33/HEH epoxide hydrolase (EPHX) mRNA, complete cds
M28882_s_at	M28882, class A, 17 probes, 17 in M28882 2907-3186, Human MUC18 glycoprotein mRNA, complete cds
U20734_s_at	U20734, class B, 20 probes, 13 in U20734cds 709-1014: 7 in reverseSequence, 7020-7258, Human transcription factor junB (junB) gene, 5 region and complete cds
S77154_s_at	S77154, class A, 20 probes, 20 in S77154 1862-2362, TINUR= NGFI-B/nur77 beta-type transcription factor homolog [human, T lymphoid cell line, PEER, mRNA, 2469 nt]
U79261_s_at	U79261, class A, 20 probes, 20 in U79261 883-1422, Human clone 23959 mRNA, partial cds.
HG2239-HT2324_r_at	Potassium Channel Protein (Gb:Z11585)
D29992_at	D29992, class C, 20 probes, 20 in all_D29992 987-1132, Human mRNA for placental protein 5 (PP5), complete cds
M27318_f_at	M27318, class A, 20 probes, 20 in M27318 365-878, Human interferon (IFN-alpha-M1) mRNA, complete cds
Z17240_at	Z17240, class C, 20 probes, 20 in all_Z17240 956-1014, Homo sapiens for mRNA encoding HMG2B

X17093_at	X17093, class C, 20 probes, 20 in all_X17093 3834-4023, Human HLA-F gene for human leukocyte antigen F
HG2147-HT2217_r_at	Mucin 3, Intestinal (Gb:M55405)
X55037_s_at	X55037, class A, 20 probes, 20 in X55037mRNA 863-1448, H.sapiens GATA-3 mRNA
M55024_s_at	M55024, class A, 20 probes, 20 in M55024 2-331, Human cell surface glycoprotein P3.58 mRNA, partial cds. /gb=M55024 /ntype=RNA
Y00787_s_at	Y00787, class C, 20 probes, 20 in all_Y00787 1314-1469, Human mRNA for MDNCF (monocyte-derived neutrophil chemotactic factor)
M30607_s_at	M30607, class A, 18 probes, 17 in M30607mRNA 2131-2301: 1 in reverseSequence, 2607, Human zinc finger protein Y-linked (ZFY) mRNA, complete cds
X70940_s_at	X70940, class B, 20 probes, 10 in X70940cds 1130-1298: 10 in reverseSequence, 1591-1722, H.sapiens mRNA for elongation factor 1 alpha-2
U72648_s_at	U72648, class B, 17 probes, 10 in U72648cds 1037-1354: 7 in reverseSequence, 4177-4210, Human alpha2-C4-adrenergic receptor gene, complete cds
M13929_s_at	M13929, class A, 20 probes, 20 in M13929mRNA 421-974, Human c-myc-P64 mRNA, initiating from promoter P0, (HLmyc2.5) partial cds
M83216_s_at	M83216, class C, 20 probes, 20 in all_M83216 3388-3584, Human aorta caldesmon mRNA, complete cds
M18700_s_at	M18700, class A, 20 probes, 20 in M18700cds 288-784, Human elastase III A gene, exon 8.
U73936_at	U73936, class A, 20 probes, 20 in U73936 5049-5523, Human Jagged 1 (HJ1) mRNA, complete cds
U12767_at	U12767, class A, 20 probes, 20 in U12767 4598-4922, Human mitogen induced nuclear orphan receptor (MINOR) mRNA, complete cds
M24283_at	M24283, class A, 20 probes, 20 in M24283mRNA 2420-2954, Human major group rhinovirus receptor (HRV) mRNA, complete cds
HG2743-HT2845_at	Caldesmon 1, Alt. Splice 3, Non-Muscle
X95735_at	X95735, class A, 20 probes, 20 in X95735 1628-2168, H.sapiens mRNA for zyxin 2
X93498_at	X93498, class A, 20 probes, 20 in X93498mRNA 589-1117, H.sapiens mRNA for 21-Glutamic Acid-Rich Protein (21-GARP)
X77909_at	X77909, class B, 20 probes, 13 in X77909cds 888-1122: 7 in reverseSequence, 1202-1406, H.sapiens IKBL mRNA
X69433_at	X69433, class C, 20 probes, 20 in all_X69433 1312-1733, H.sapiens mRNA for mitochondrial isocitrate dehydrogenase (NADP+)
X66401_cds1_at	X66401, class B, 20 probes, 11 in X66401cds#1 327-615: 6 in fullSequence, 45931-47208: 3 not in GB record, LMP2 gene extracted from H.sapiens genes TAP1, TAP2, LMP2, LMP7 and DOB
X66114_rna1_at	X66114, class A, 20 probes, 20 in X66114mRNA 564-1074, H.sapiens gene for 2-oxoglutarate carrier protein
X54232_at	X54232, class A, 20 probes, 20 in X54232mRNA 3259-3643, Human mRNA for heparan sulfate proteaglycan (glypican)
X16832_at	X16832, class C, 20 probes, 20 in all_X16832 840-1381, Human mRNA for cathepsin H (EC 3.4.22.16)
X06562_at	X06562, class C, 20 probes, 20 in all_X06562 3951-4396, Human mRNA for growth hormone receptor
X06482_at	X06482, class A, 20 probes, 18 in X06482cds 60-405: 2 in reverseSequence, 884-887, Human theta 1-globin gene
V01512_rna1_at	V01512, class A, 20 probes, 20 in V01512mRNA#2 1533-2061, Human cellular oncogene c-fos (complete sequence)
U60115_at	U60115, class A, 20 probes, 20 in U60115 1863-2211, Human skeletal muscle LIM-protein SLIM1 mRNA, complete cds
U60062_at	U60062, class A, 20 probes, 20 in U60062 1060-1550, Human FEZ1-T mRNA, alternatively spliced form, complete cds
U52111_rna3_at	U52111, class A, 20 probes, 20 in U52111mRNA#3 2176-2659, Homo sapiens Xq28 genomic DNA in the region of the ALD locus containing the genes for creatine transporter (SLC6A8), CDM, adrenoleukodystrophy (ALD), Na+-isocitrate dehydrogenase gamma subunit (IDH
U49928_at	U49928, class A, 20 probes, 20 in U49928 2513-3035, Human TAK1 binding protein 1 (TAB1) mRNA, complete cds
U47926_at	U47926, class A, 20 probes, 20 in U47926 1546-1996, Human unknown protein B mRNA, complete cds
U18018_at	U18018, class A, 20 probes, 20 in U18018 1732-2290, Human E1A enhancer binding protein (E1A-F) mRNA, partial cds
U16954_at	U16954, class A, 20 probes, 20 in U16954 1099-1579, Human (AF1q) mRNA, complete cds
U13220_at	U13220, class A, 20 probes, 20 in U13220 1586-2066, Human forkhead protein FREAC-2 mRNA, partial cds
U09646_at	U09646, class A, 20 probes, 20 in U09646exon 358-874, Human carnitine palmitoyltransferase II precursor (CPT1) gene

U09414_at	U09414, class A, 20 probes, 20 in U09414 1994-2462, Human zinc finger protein ZNF137 mRNA, complete cds
U03911_at	U03911, class A, 20 probes, 20 in U03911 2485-3013, Human mutator gene (hMSH2) mRNA, complete cds
U03398_at	U03398, class A, 20 probes, 20 in U03398 1069-1576, Human receptor 4-1BB ligand mRNA, complete cds
S78085_at	S78085, class A, 20 probes, 20 in S78085 719-1187, PDCD2=programmed cell death-2/Rp8 homolog [human, fetal lung, mRNA, 1282 nt]
M92934_at	M92934, class A, 20 probes, 20 in M92934mRNA 1492-2026, Human connective tissue growth factor, complete cds
M69225_at	M69225, class A, 20 probes, 20 in M69225mRNA 8371-8845, Human bullous pemphigoid antigen (BPAG1) mRNA, complete cds
M64098_at	M64098, class A, 20 probes, 20 in M64098 3873-4305, Human high density lipoprotein binding protein (HBP) mRNA, complete cds
M57567_at	M57567, class A, 20 probes, 20 in M57567 491-953, Human ADP-ribosylation factor (hARF5) mRNA, complete cds
M29971_at	M29971, class A, 20 probes, 20 in M29971 282-750, Human 6-O-methylguanine-DNA methyltransferase (MGMT) mRNA, complete cds
M22898_at	M22898, class A, 20 probes, 20 in M22898mRNA 2042-2600, Human phosphoprotein p53 gene
L42450_at	L42450, class A, 20 probes, 20 in L42450mRNA 1022-1448, Homo sapiens pyruvate dehydrogenase kinase isoenzyme 1 (PDK1) mRNA, complete cds
L38487_at	L38487, class A, 20 probes, 20 in L38487mRNA 1623-2115, Human estrogen receptor-related protein (hERRa1) mRNA, 3 end, partial cds
L20971_at	L20971, class A, 20 probes, 20 in L20971 3698-3992, Human phosphodiesterase mRNA, complete cds
L13197_at	L13197, class A, 20 probes, 20 in L13197 1853-2099, Human (clone D21S418E) pregnancy-associated plasma protein A (PAPP-A) gene, 5 UTR
K03008_cds2_at	K03008, class C, 20 probes, 2 in K03008cds 90-118: 18 not in GB record, gamma-G2-psi gene extracted from Human gamma-C-crystallin (gamma-3) gene, gamma-G2-psi gene extracted from Human gamma-C-crystallin (gamma-3) gene
J02854_at	J02854, class A, 20 probes, 20 in J02854 531-1089, Human 20-kDa myosin light chain (MLC-2) mRNA, complete cds
HG3123-HT3299_at	Homeotic Protein Gbx2
HG2663-HT2759_at	Homeotic Protein Emx2
HG1612-HT1612_at	Macmarcks
D87458_at	D87458, class A, 20 probes, 20 in D87458 3244-3784, Human mRNA for KIAA0282 gene, partial cds
D82346_at	D82346, class A, 20 probes, 20 in D82346 944-1316, Human mRNA for HNSPC, complete cds
D63482_at	D63482, class A, 20 probes, 20 in D63482 1722-2226, Human mRNA for KIAA0148 gene, complete cds
D49490_at	D49490, class A, 20 probes, 20 in D49490 1092-1644, Human mRNA for protein disulfide isomerase-related protein (PDIR), complete cds
D31764_at	D31764, class A, 20 probes, 20 in D31764 1478-1982, Human mRNA for KIAA0064 gene, complete cds
D21267_at	D21267, class A, 20 probes, 20 in D21267mRNA 1481-1979, Human mRNA for highly expressed protein