Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Scienze Farmacologiche e Tossicologiche,
dello Sviluppo e del Movimento Umano

Ciclo XXVIII

Settore Concorsuale di afferenza: 05/F1 - BIOLOGIA APPLICATA

Settore Scientifico disciplinare: BIO/13 - BIOLOGIA APPLICATA

Genomic and post-genomic analysis of human chromosome 21 in relation to the pathogenesis of trisomy 21 (Down syndrome)

Presentata da: Maria Caracausi

Coordinatore Dottorato **Relatore**

Prof.ssa Patrizia Hrelia **Prof. Pierluigi Strippoli**

Esame finale anno 2016

Abstract

We performed an innovative systematic meta-analysis of gene expression profiles of whole normal human brain and heart to provide a quantitative transcriptome reference map of it, i.e. a typical reference value of expression for all the known, mapped and uncharacterized (unmapped) transcripts. For this reason, we used the software named TRAM (Transcriptome Mapper), which is able to generate transcriptome maps based on gene expression data from multiple sources. We also analyzed differential gene expression by comparing brain with human foetal brain, with a pool of non-brain tissues and with the three brain sub-region: cerebellum, cerebral cortex and hippocampus, the main regions severely affected with cognitive impairment, as seen in the case of trisomy 21. Data were downloaded from microarray databases, processed and analyzed using TRAM software and validated *in vitro* by assaying gene expression through several magnitude orders by "Real Time" reverse transcription polymerase chain reaction (RT-PCR). The excellent agreement between *in silico* and experimental data suggested that our transcriptome maps may be a useful quantitative reference benchmark for gene expression studies related to the human brain and heart.

We also generated an integrated quantitative transcriptome map by systematic meta-analysis from all available gene expression profile datasets related to AMKL in paediatric age. The incidence of Acute Megakaryoblastic Leukemia (AMKL) is 500-fold higher in children with Down Syndrome (DS) compared with non-DS children. We present an integrated original model of the DS AMLK transcriptome, providing the identification of genes relevant for its pathophysiology which can potentially be new clinical markers.

Finally, computational and molecular analysis of a highly restricted region of chromosome 21, which represents a strong candidate for typical DS features and is considered as intergenic, was performed. Northern Blot analysis and computational biology results show that HR-DSCR contain active loci bidirectionally transcribed.

Key words: Human Brain; Human Heart; Gene Expression Profile; Integrated Transcriptome Reference Map; Meta-analysis; Human chromosome 21; Down Syndrome (Trisomy 21); Acute Megakaryoblastic Leukemia (AMKL); Down syndrome critical region (DSCR); Intellectual disability.

Contents

Chapter 1

Introduction

1.1 Down syndrome

Down Syndrome (DS) is the most frequent human chromosomal disorder, with a frequency of 1 in ~400 conceptions and 1 in ~700 births (Morris et al., 1999; Parker et al., 2010). Main symptoms include intellectual disability (ID), cardiovascular defects and craniofacial dysmorphisms (Gardiner, 2010). More in detail, these symptoms may be observed: distinct facial and physical features among which are almond shaped eyes (due to epicanthal folds), a small, somewhat flat nose, a small mouth with a protruding tongue, a single crease across the palms, inturned fifth finger of their hands, a larger than normal space between the first and second toes; language production deficiency; cognitive impairment, which is present to some degree of severity in all affected individuals and which involves symbolic thought, whereas affectivity and social skills are conserved; cardiac defects in about 40% of cases; Hirschprung's disease; hypotonia; visual and hearing impairments; increased risk of leukemia, in particular megakaryoblastic; immune disorders, including increased susceptibility to infections and to autoimmune pathologies such as alopecia and celiac disease (CD); endocrine disorders, including hypothyroidism; and early-onset cognitive decline with neuropathological alterations similar to those observed in the brains of patients with Alzheimer's disease (AD) (Epstein, 1989; Roizen and Patterson, 2003; Mégarbané et al., 2009; Gardiner, 2010; Letourneau and Antonarakis, 2012; Hickey et al., 2012).

In 1959, the young French doctor Jérôme Lejeune (1926-1994) published, together with Marthe Gautier and Raymond Turpin, the finding of an additional chromosome 21 in nine children with DS (Lejeune et al., 1959). This condition has been called trisomy 21, a genetic mutation leading to the presence of three copies of human chromosome 21 (Hsa21), instead of the normal two, in the cells of the affected individuals. This discovery is commonly recognized as a milestone in the history of genetics (National Human Genome Research Institute 2013), because it introduced the notion that a given clinical symptom may be connected to a specific alteration of the human genetic material for the first time, giving origin to the field of medical genetics. The DS phenotype is expected to be associated with an altered expression of the genes located on Hsa21 (Sinet et al., 1975; Gardiner and Costa, 2006; Roper and Reeves, 2006; Pritchard et al., 2008; Korenberg, 2009; Patterson, 2009). Although DS was the first genetic alteration to have been described in humans and the most frequent form of ID caused by a microscopically demonstrable chromosomal aberration, its molecular pathogenesis is still unknown.

There has been much debate about the existence of selected, critical regions or genes on Hsa21 as the main one responsible for ID or for the other symptoms (Rahmani et al., 1990;

Korenberg et al., 1990; Delabar et al., 1993). Since the first reports in the '70's (e.g. Raoul et al., 1976 from the group of Lejeune), descriptions of partial (segmental) trisomies have accumulated, although at slow pace. Two landmark studies in 2009, each analyzing about 30 cases of these conditions, have concluded that it is not likely that there is a single region candidate for DS phenotype (Down Syndrome Candidate/Critical Region, DSCR) on Hsa21 (Lyle et al., 2009; Korbel et al., 2009). However, it remains to be explained how such a homogeneous phenotype, though with its natural individual variability, may be generated by completely different portions of the same chromosome when involvement of a single different gene is usually associated with completely different phenotypes of monofactorial origin. We believe that this is still a fundamental line of research. In fact, even if it would not be possible to affirm with certainty that a gene, if present in an extra copy, is more responsible than the other for a symptom, it can always be excluded that a gene not present in three copies in at least one subject with a typical DS phenotype is critical for ID, the nearly universal symptom in DS. On the other hand, subjects with nearly absent ID and partial trisomy 21 might offer complementary clues to critical regions. It is important to critically evaluate all described cases and to continue searching for partial trisomy proposing molecular cytogenetics (CGH-Array) investigations (Melis et al., 2011) in particular in cases with a discordant genotype-phenotype relationship.

1.2 Gene expression profiling

Firm conclusions about the role of a DNA element based on its chromosomal localization can be derived only if the annotated physical map of the chromosome is actually complete and accurate as much as possible. Gene expression analysis performed by TRAM software (Lenzi et al., 2011) shows that there are hundreds of uncharacterized transcripts on Hsa21 (preliminary data). Although continuing the patient reconstruction of each single locus structure, expression and function appears to have fallen out of research trends in the last 10 years, without a really complete list of Hsa21 elements we could miss the identification of a critical component on an interesting Hsa21 region because it is not highlighted on the map (e.g., ENCODE Project Consortium et al., 2012). As it was shown by identifying a large novel gene extending for more than 100,000 bp on Hsa21, encoding a protein (CYYR1) conserved in vertebrates (Vitale et al., 2002; Vitale et al., 2007) and having gone unrecognized in the putatively complete gene catalogue published with the report of the whole Hsa21 sequence (Hattori et al., 2000), due to plasticity of gene expression the "manual" characterization of individual genomic regions is an ever-going process that should deserve active attention even in the post-genomic era.

The most complex organ for gene expression profile investigation is the brain (Oldham et al., 2008). The brain is the main organ of the Central Nervous System (CNS). It is an immensely complex organ composed of billions of precisely interconnected neurons which together allow it to carry out sensory, motor and cognitive functions. Impaired development of cognitive functions is the cause of intellectual disability (ID). The causes of ID may be genetic or non-genetic: they are of genetic origin if they are due to abnormalities in a single gene or to structural or numerical abnormalities of chromosomes. Trisomy 21 (Down syndrome) is the best known genetic cause of ID (Mégarbané et al., 2013). *Homo sapiens* is the mammal with the highest number of genes expressed in the brain (Lockhart and Barlow, 2001; Enard et al., 2002). Numerous studies show that children with DS exhibit cerebral developmental delay beginning in the first year of life, including reductions in dendritic ramification, synaptogenesis and neuronal cell numbers. Specific reductions in the number of principle neurons affect the hippocampus, in particular, the pyramidal cells of the Ammon's horn 1 and 3 sub-regions (CA1 and CA3) (Haydar and Reeves, 2012). The majority of studies have shown that these hippocampal abnormalities affect the long-term storage of explicit memories in DS patients, causing syndrome specific symptoms which are more severe in individuals with DS than with other mental retardation conditions (Carlesimo et al., 1997; Dierssen et al., 2009).

The altered gene dosage in trisomy 21 also increases the susceptibility to congenital heart defects in DS individuals (Ramachandran et al., 2015). Several hereditary cardiovascular malformations syndromes are caused by gene products implied in the disease processes and in the pathways regulating cardiac development (Vaughan and Basson, 2000). Common cardiovascular malformations are, for example, atrial (ASDs) and ventricular septal defects. Atrioventricular septal defects (AVSDs) are uncommon cardiovascular malformations in the general population (Hartman et al., 2011) but they are a common type of congenital heart defect in Down syndrome individuals. Much progress has been made about the cardiovascular malformation repair since the first successful classification of the atrioventricular canal and the surgical procedure developed by Rastelli and coll. (Rastelli et al., 1966; Rastelli et al., 1968).

Transcriptome studies are used to understand physiologic mechanisms and pathways or genes involved in many biologic processes (Caracausi et al., 2014). Alteration of the gene expression level could help to identify alterations in cell function. Transcriptome is a promising approach for diagnostic evaluation in routine clinical practice and could be useful to predict disease development and progression.

RNA microarrays and RNA sequencing (RNA-Seq) are considered the two main kinds of high-throughput technologies employed to assess gene expression studies (Costa and Franco, 2015). Recently, several aspects of data obtained from microarrays and RNA-seq have been

compared in a number of published studies (reviewed by Wang et al., 2014). RNA-seq was found to be superior to microarray due to a higher sensitivity in detecting low abundance transcripts (Wang et al., 2014; Zhao et al., 2014) as well as to a broader dynamic range (Zhao et al., 2014), and it also allows avoidance of cross-hybridization (Zhao et al., 2014) and detection of new transcripts (Guo et al., 2013). Despite the benefits of RNA-Seq, microarrays remain an accurate tool for measuring the levels of gene expression (Malone and Oliver, 2011), and gene expression-based predictive models generated from RNA-seq and microarray data were found to be similar (Wang et al., 2014). Overall, in large comparative studies, these two methods have produced comparable results in terms of gene expression profiling (Guo et al., 2013; Zhao et al., 2014). Microarrays still appear to be the most common choice of researchers when conducting transcriptional profiling experiments. This is probably because: RNA-Seq sequencing technology is new to most researchers, it is still more expensive than microarray, and data storage and analysis are more complex (Zhao et al., 2014).

The huge amount of transcriptomic studies performed with microarray technology, stored in publicly available databases, still provide excellent data mining resources for researchers (Guo et al., 2013). There have been several microarray experiments performed on the whole human brain and its sub-regions, and on the whole human heart, to analyze the gene expression profile of the whole organs, to compare gene expression patterns between different brain sub-regions (Khaitovich et al., 2004; Roth et al., 2006) and to compare pathophysiological states (Iwamoto et al., 2004; Hodges et al., 2006).

Our goal in this study was to perform a systematic meta-analysis of the gene expression profile of the whole normal human brain and of its sub-region, and of the whole normal human heart in order to provide a quantitative transcriptome reference map of them, i.e. a reference typical value of expression for each of the known, mapped and uncharacterized (unmapped) transcripts assayed by any of the experimental platforms used to this regard. This task implied the possibility of performing cross-platform, inter-array and intra-array data normalization in order to incorporate any publicly available dataset in the calculation. To this aim, we used the software named TRAM (Transcriptome Mapper), which is able to generate transcriptome maps of this type from any source listing gene expression values for a given gene set, e.g. expression microarray (Lenzi et al., 2011). In addition to providing reference gene expression values (in the form of percentages of the mean value), allowing quantitative comparisons among the expression of all investigated genes, it maps gene expression along the chromosome's physical map, thus also allowing discovery of regional over- or under-expression within a biological condition or while comparing two different biological conditions.

The European Bioinformatics Institute (EBI) and National Center for Biotechnology Information (NCBI) provide Array Express (Brooksbank et al., 2014) and GEO (Gene Expression Omnibus) (Barrett and Edgar, 2006) databases respectively, which are repositories of high throughput sequencing studies and hybridization array data that can be searched for and downloaded.

The use of whole (*in toto*) brain is justified by its recognition as a specific organ performing superior cognitive functions, and a wealth of gene expression data has been produced from whole brain RNA by different methods (among many others, for example: Northern Blot analysis, expressed sequence tag (EST) libraries, polymerase chain reaction (PCR) assays and microarray experiments). Due to the very recent demonstration of differences in adult brain regional gene expression according to sex (Trabzuni et al., 2013), we extended our analysis of the whole brain to the comparison of male- vs. female-derived samples.

In addition, we extended our analysis to a comparison with human foetal brain gene expression, with a pool of human tissues and to three normal human brain regions: cerebellum, cerebral cortex and hippocampus: three of the main brain regions severely affected when cognitive impairment occurs (Kesslak et al., 1994; Nadel, 2003; Pennington et al., 2003), as happens in the case of trisomy 21 (Haydar and Reeves, 2012).

This also yielded biological insights about those genes which have an intrinsic over/under-expression in the brain or brain subregions, thereby offering a basis for the regional (chromosomal) analysis of gene expression. This could be useful for the study of chromosomal alterations associated to cognitive impairment, such as trisomy 21, the most common genetic cause of ID.

To date, human studies of gene expression profile in cardiac muscle are limited. Those available are often not comparable, not comprehensive or not related to the position of each gene on the genome map. The whole heart transcriptome map could be used to infer a very high rate of information about the healthy cardiac muscle, and could be used to analyse the differential transcription profile respect to a different biological condition, as for example the diseased myocardium.

Finally, data obtained by our meta-analysis were validated by assaying gene expression through several magnitude orders by "Real Time" reverse transcription polymerase chain reaction (RT-PCR). We found an excellent agreement between *in silico* and experimental data, thus suggesting that our transcriptome maps may be a useful tools as a quantitative references for gene expression studies related to these human tissues.

Basic research on DS is now rapidly accelerating, and there is the possibility that the results will be beneficial for individuals with DS (Antonarakis and Epstein, 2006). Several

studies have shown that individuals with DS have a specific cancer risk pattern, or tumor profile: their risk of developing leukemia and testicular cancer is much higher than age-matched controls, while women with DS almost never develop breast cancers (Hasle et al., 2000; Patja et al., 2006). In particular, children with DS show an increased prevalence of acute leukemia, both lymphoid (ALL) and myeloid (AML), with relative risk ranging from 10 to 20 times higher than the normal population (Hitzler and Zipursky, 2005; Massey, 2005). In nearly half of the cases, these childhood leukemias are classified as megakaryoblastic leukemia (AMKL), a relatively rare subtype of AML also known as AML M7, according to French–American–British (FAB) classification, whose incidence increases by 500-fold in children with DS by the age of 4 years as compared to the chromosomally normal population (reviewed in (Khan et al., 2011). This observation strongly suggests that trisomy 21 directly contributes to the neoplastic transformation of hematopoietic cells, in particular in the megakaryocyte lineage cells. Interestingly, acute leukemia cells harboring megakaryocyte markers and presenting in subjects without DS may show trisomy 21 (Paolini et al., 2003). We also described a cell line derived from blast cells of a patient with type M2 AML wich has trisomy 21 and megakaryocyte features (Bonsi et al., 1997). More recently, mutations of the gene encoding for the transcription factor GATA1 have been shown to cooperate with trisomy 21 in initiating megakaryoblastic proliferation in nearly all DS AMKL cases while they are absent in non-DS AMKL (Wechsler et al., 2002; Khan et al., 2011). GATA1 mutations in DS cells give rise to a short, truncated form of GATA1 (GATA1s) transcription factor that, in this form, is not able to establish normal interactions with other gene regulators (Klusmann et al., 2010a).

Transient myeloproliferative disorder (TMD) is a clonal pre-leukemia condition, occurring in 10% of children with DS during the neonatal period, presenting at a median age of 3-7 days with accumulation of immature megakaryoblasts (Khan et al., 2011). TMD cases usually resolve spontaneosuly, but DS AMKL may develop within 1-4 years in 20-30% of these children. AMKL may develop in non-DS children, usually at an higher age in comparison to DS subjects (median 8 vs. 1.8 years, respectively) and in absence of a trisomy 21 background. Cytogenetic abormalities described in non-DS AMKL cells include trisomy 8 and 1 and monosomy 7 (Khan et al., 2011).

An open issue is the relevance of trisomy 21 as a specific background for the higher incidence of AMKL in DS. A few previous studies have used gene expression profiling by microarray analysis in order to identify specific transcriptome alterations in DS and/or non-DS AMKL, as well as in TMD (Klusmann et al., 2010a; Klusmann et al., 2010b; Yagi et al., 2003; Lightfoot et al., 2004; McElwaine et al., 2004; Bourquin et al., 2006; Ge et al., 2006; Radtke et al., 2009). Due to the rarity of AMKL, these works often analyze a small number of cases, using

a variety of experimental platforms. Results were consequently affected by a small grade of comparability.

One of the first goals of this work was to perform a systematic meta-analysis using any available gene expression profile dataset related to AMKL in pediatric age in order to produce a differential transcriptome map between DS and non-DS related AMKL. For the generation and the analysis of quantitative transcriptome maps we used TRAM (Transcriptome Mapper) (Lenzi et al., 2011). The comparison of 43 DS-AMKL samples with 45 non-DS AMKL samples represents the largest study on the subject, highlighting the relevance of trisomy 21 in the development of AMKL in comparison with AMKL originating from non-trisomic cells. Results show significant over- or under-expression of distinct chromosomal segments and of single key genes in the whole genome, as well as on chr21, adding new knowledge compared with that produced by the single works from which the data were originally obtained. In addition, each considered type of leukemia was compared with the expression profile of TMD cells and normal human megakaryoblast/megakaryocyte cells (MK), allowing the building of a model for the disorder in differentiation process that lead to DS and non-DS AMKL. Comparisons with cord blood-derived MK cells (CB MK) have also been performed, due to the fact that leukemias in infants or young children originate from fetal hematopoietic cells (Klusmann et al., 2010a; Klusmann et al., 2010b; Chou et al., 2008; Tunstall-Pedoe et al., 2008) and the progenitor cells (fetal/neonatal MKP) are present in the cord blood (CB) (Olson et al., 1992; Liu and Sola-Visner, 2011). For each cell type investigated, reference expression data for about 17,000-26,000 mapped sequences have been generated and validated through a sample comparison with known data. The biological and clinical significance of these data is discussed.

1.3 HR-DSCR on human chromosome 21 and gene hunting

A "Down Syndrome critical region" (DSCR) sufficient to induce the most constant phenotypes of Down syndrome (DS) had been identified by studying partial (segmental) trisomy 21 (PT21) as an interval of 0.6-8.3 Mb within human chromosome 21 (Hsa21), although its existence was later questioned. In the work of Pelleri and coll. (2016) an innovative, systematic reanalysis of all described PT21 cases (from 1973 to 2015) was performed. In particular, an integrated, comparative map from 126 cases with or without DS fulfilling stringent cytogenetic and clinical criteria was built (Pelleri et al., 2016). The map allowed us to define or exclude fine Hsa21 sequence intervals as candidates for DS, also integrating duplication copy number variants (CNVs) data. A highly restricted DSCR (HR-DSCR) of only 34 kb (kilobase) on distal 21q22.13 has been identified as the minimal region whose duplication is shared by all DS

subjects and is absent in all non-DS subjects. Also being spared by any duplication CNV in healthy subjects, HR-DSCR represents a strong candidate for the typical DS features, the intellectual disability and some facial phenotypes. HR-DSCR contains no known gene and has homology only to the chimpanzee genome. Searching for HR-DSCR functional loci should become a priority for understanding the fundamental genotype-phenotype relationships in DS.

We performed a systematic analysis of computational and molecular biology on HR-DSCR, a region known as an intergenic region, but suspected by the study of partial trisomy 21 as a potential host of active loci involved in critical functions for the manifestation of the intellectual disabilities in DS. This analysis was performed to test the hypothesis that it is not an intergenic region, but rather expressing transcripts not yet identified. For this reason, we applied systematically advanced tools of gene identification by computational biology (Blastn, BLASTX, FGENESH, miRBase), verifying, through different methods of molecular biology, the validity of the predictions *in vitro* using human tissues (Northern Blot, polymerase chain reaction (PCR), sequencing automatic). The results show that the systematic analysis of the region leads to the classification of a critical region of human chromosome 21 as a new active locus transcribed in both directions. This locus is formally identified and an initial characterization of its structure and expression is also presented.

Chapter 2

Aim of the thesis

The research project has the objective to study and analyse the human chromosome 21 genes to understand the molecular mechanism of DS and to find therapeutic approaches for the cure of the syndrome. The cause of the syndrome is a genetic mutation that results in the presence of chromosome 21 in 3 copies in the cells of the affected individuals. The typical symptoms and signs are attributed to altered expression of genes located on chromosome 21 and include a "facies" feature; intellectual disabilities; heart defects; muscle hypotonia; increased risk of leukemia, in particular megakaryoblastic; and early biological aging in some cases similar to that of Alzheimer's disease. The molecular mechanisms of the syndrome are still unknown and it is difficult to identify specific chromosome 21 genes responsible for different symptoms.

Full identification and characterization of these genes is not yet complete; currently about 420 genes and gene models on the long arm of chromosome 21, and 4 on the short arm of chromosome 21 are described, but there is information available for only 145 of them (Patterson, 2009). Basic research on trisomy 21 could propose possible therapeutic approaches for DS (Antonarakis and Epstein, 2009).

The research conducted during the doctoral period enabled us to create and analyze tissue specific gene activity maps of normal human tissues primarily involved in DS and to make comparisons of gene expression between normal and trisomic cells. In particular, we have created transcriptome maps of the whole normal human brain and of three brain sub-regions (cerebellum, cerebral cortex and hippocampus), of the whole heart, and of blood cells, searching for all data available in the literature and in the databases of gene expression profiles performed on microarray platforms and integrating these data using the TRAM software (mapper transcriptome) (Lenzi et al., 2011). We focused on these organs and tissues because 100% of individuals with DS have intellectual disabilities, and the cerebellum, cerebral cortex and hippocampus are the three sub-regions severely affected when ID occurs, 30-40% of DS individuals suffer from congenital heart defects, and furthermore because the risk to have acute megakaryoblastic leukemia (AMKL) is 500-fold higher in DS compared with non-DS children.

These maps provide a typical reference value of expression for each of the known, mapped and uncharacterized loci of the whole human tissues and organs cited above. These global portraits of gene expression could be used to test hypotheses about localized gene expression levels of human transcripts and could also contribute to a better understanding on a regional (chromosomal) basis of the chr21 genes expression. The comparison between normal and trisomy 21 cells was done for the blood cells in particular between AMKL and MK (megakaryocytes) in conditions and not of trisomy 21 in order to identify new molecular markers for the progression to AMKL in DS. These analyses allowed us to confirm that chr21 genes have an average of expression higher than the genes of other chromosomes both in DS and non-DS

AMKL cells, and to identify possible markers of progression from normal MK cells to neoplastic cells.

Another objective was to perform computational and molecular biology analysis of a highly restricted region of chr21, critical for DS (HR-DSCR) located within the larger critical region for DS (DSCR) (unpublished preliminary data). This region has been identified by a systematic study of 127 cases of partial trisomy 21, in which different segments of chromosome 21 associated or not to the cardinal signs of the syndrome are present in three copies. This region of only 34 kb on distal 21q22.13 is the minimal region whose duplication is shared by all DS subjects and is absent in all non-DS subjects.

To date the HR-DSCR was defined as an intergenic region, but it represents a strong candidate for the typical DS features, the intellectual disability and some facial phenotypes. The analysis was performed to demonstrate that it potentially hosts active loci involved in critical functions which causes DS symptoms. Searching for HR-DSCR functional loci should become a priority for understanding the fundamental genotype-phenotype relationships in DS.

All these goals converge to elaborate on theories of the functioning of chromosome 21, to build an overall pathogenetic model for the symptoms of trisomy 21 and to identify new therapeutic approaches.

Chapter 3

Materials and Methods

3.1 Systematic meta-analysis of gene expression profiles of normal human tissues

3.1.1 Gene expression database search

In order to retrieve datasets derived from whole normal adult human brain, foetal brain, cerebellum, cerebral cortex, hippocampus and whole normal adult heart, we made a systematic search in gene expression data repositories for any single sample available listing gene expression values for these tissues. Gene Expression Omnibus (GEO) functional genomics repository was searched for: "Homo sapiens [ORGANISM] AND brain" (or "cerebellum", or "cortex", or "hippocampus", or "heart"). ArrayExpress database of functional genomics experiments was searched at http://www.ebi.ac.uk/arrayexpress/ for the terms "brain" (or "cerebellum", or "cortex", or "hippocampus", or "heart"), choosing "Homo sapiens" as organism. This strategy was used to ensure high sensitivity in the search. Search results were then filtered using inclusion and exclusion criteria as explained below in the Dataset selection section.

In addition, the term "Tissue" was used to retrieve datasets derived from collection of different human tissues analyzed in the same databases (in GEO: "Homo sapiens [ORGANISM] AND tissue*[TI] OR organ*[TI]"; in ArrayExpress: "Tissue", choosing "Homo sapiens" as organism). This led to generate a pool of samples including all the main human organs and tissues to generate a comparison set to highlight brain-specific or heart-specific differential gene expression compared to all the other anatomical human structures.

The searches were made up to May 2013.

3.1.2 Gene expression dataset selection

The inclusion criteria of datasets in the analysis were: experiments carried out on the whole organ or tissue; normal phenotype of individuals; adult or foetal (for the foetal brain transcriptome map only) age of the subject from whom the sample was obtained; availability of the raw or pre-processed data.

Exclusion criteria were: exon arrays (hampering data elaboration by TRAM due to an exceedingly high number of data rows) or platforms using probes split into several distinct arrays for each sample (hampering intra-sample normalization); lack of identifiers corresponding to those found in the GEO Sample records (GSM) or Array Express sample records; platforms assaying an atypical number of genes (i.e. <5.000 or >60.000); data derived from cell lines, pathological or treated tissue, children or foetal (except in foetal brain transcriptome map) tissues.

In order to obtain a quantitative transcriptome map, values from each dataset were linearized when provided as logarithms. In some cases we used raw files (e.g. File CEL) to be converted into pre-processed data, using the software "Alt Analyze" (Emig et al., 2010).

3.1.3 TRAM (Transcriptome Mapper) analysis

TRAM (Transcriptome Mapper) software (Lenzi et al., 2011) allows the import of gene expression data recorded in the NCBI GEO (Gene Expression Omnibus) and EBI Array Express databases in tab-delimited text format. It also allows the integration of all data by decoding probe set identifiers to gene symbols via UniGene data parsing (Lenzi et al., 2006), normalizing data from multiple platforms using intra-sample and inter-sample normalization (scaled quantile normalization) (Piovesan et al., 2013), creating graphical representation of gene expression profile through two ways, "Map" and "Cluster" mode, and determining the statistical significance of results.

We created a directory (folder) for each tissue, containing all the sample datasets related to the same source and selected for the study. To compare brain samples with a pool of human tissues (without brain samples) we collected the first in a folder named Pool 'A' and the second in a folder named Pool 'B' (Table 1); to compare foetal brain with adult total brain we collected the first in a folder named Pool 'C (Table 1)'; to compare the brain regions with total brain we collected the samples from cerebellum in a folder named Pool 'D', the samples from cerebral cortex in a folder named Pool 'E', the samples from hippocampus in a folder named Pool 'F' (Table 1). In addition, we also provided two datasets deriving from the adul brain samples for which the sex of the sample donor was available: male brain (Pool 'A.1') and female brain (Pool 'A.2') (Table 1). We performed the same thing for the hippocampus, providing two datasets deriving from the hippocampus samples for which the sex of the sample donor was available: male hippocampus (Pool 'F.1') and female hippocampus (Pool 'F.2') (Table 1). We did not consider samples deriving from male/female mix of tissue. To compare whole heart samples with a pool of non-cardiac tissues, we collected the first in a folder named Pool 'G' and the second in a folder named Pool 'H' (Table 1).

Pool	Sample type	Sample number	TRAM mapped loci
Pool 'A'	Whole adult brain	n=60	39,250
Pool 'A.1'	Male adult brain	n=13	27,437
Pool 'A.2'	Female adult brain	n=5	27,954
Pool 'B'	Pool of non-brain tissue	n=622	34,985
Pool 'C'	Foetal brain	n=35	38,482
Pool 'D'	Cerebellum	n=140	38,163
Pool 'E'	Cerebral cortex	n=18	27,504
Pool 'F'	Hippocampus	n=41	30,739
Pool 'F.1'	Male hippocampus	n=15	26,045
Pool 'F.2'	Female hippocampus	n=14	26,045
Pool 'G'	Whole heart	n=32	43,360
Pool 'H'	Pool of non-cardiac tissue	n=629	35,000

Table 1. Sample pools selected for the meta-analysis of gene expression profiles in adult brain (pool 'A'), male adult brain (pool 'A.1'), female adult brain (pool 'A.2'), pool of non-brain tissues (pool 'B'), foetal brain (pool 'C'), cerebellum (pool 'D'), cerebral cortex (pool 'E'), hippocampus (pool 'F'), male hippocampus (pool 'F.1'), female hippocampus (pool 'F.2'), whole hart (pool 'G'), pool of non-cardiac tissue (pool 'H').

The comparisons allowed the analysis of differential transcriptome maps, using the ratio of the mean expression values for each locus, in addition to the maps related to each single type of sample.

We ran the whole set of analyses permitted by TRAM (in both "Map" and "Cluster" mode, although we focused on the "Map" mode) using default parameters as described (Lenzi et al., 2011). We used an updated version of TRAM including enhanced resolution of gene identifiers and updated UniGene and Entrez Gene databases (TRAM 1.1, June 2013), in comparison with the original 2011 version (Lenzi et al., 2011). TRAM is freely available at http://apollo11.isto.unibo.it/software. Briefly, gene expression values were assigned to individual loci via UniGene, intra-sample normalized as percentage of the mean value and inter-sample normalized by scaled quantile. The value for each locus, in each biological condition, is the mean value of all available values for that locus. The whole genome gene expression median value was used in order to determine percentiles of expression for each gene. Although TRAM is a map-centred transcriptome analysis tool it can also summarize and allow the analysis of gene expression data of unmapped genes, exploiting its capability of parsing and normalization in order to highlight differential expression of single genes between two biological conditions even in the absence of data about genomic location of the gene (Lenzi et al., 2011).

Using the "Map" mode graphical representation we searched for over/under-expressed genome segments, which have a window size of 500,000 bp and a shift of 250,000 bp. The expression value for each genomic segment is the mean of the expression values of the loci included in that segment. A segment is defined over/under-expressed if it has an expression value which is significantly different between two conditions analyzed, and contains at least 3

individually over/under-expressed genes, e.g. genes which have expression values within the highest and the lowest 2.5th percentile. Significance of the over/under-expression for single genes was determined by running TRAM in "Map" mode with a segment window of 12,500 bp. This window size corresponds to about a fifth of the mean length of a gene, so the significant over/under-expression of a segment almost always corresponds with that of a gene. When the segment window contains more than one gene, the significance is maintained if the expression value of the over/under-expressed gene prevails over the others.

For the creation of the maps, TRAM software does not consider probes where the expression values is not available, assuming that an expression level has not been measured. Furthermore, it gives 95% of the minimum positive value present in a sample to those expression values equal to or lower than "0", in order to obtain meaningful numbers when we need to obtain a ratio between values in pool 'A' and pool 'B'. Assuming that in these cases an expression level is too low to be detected under the experimental conditions used, this transformation is useful to highlight differential gene expression.

3.1.4 Housekeeping gene search

We determined the predicted genes that behave like housekeeping genes, in that they are mainly involved in fundamental cellular function and are universally and constantly expressed in all tissues (Butte et al., 2001; Tu et al., 2006). A search of housekeeping genes in the transcriptome maps was performed using the following parameters in combination: expression value >100, in order to select genes expressed above the mean value and so at an appreciable level; data points number ≥ of half the number of samples of the map, in order to select commonly expressed genes; SD, expressed as a percentage of the mean value, ≤30 or ≤40, in order to identify genes with a low expression variation among different samples. We searched for SD ≤30 to identify the first gene with the lowest SD, instead SD ≤40 to identify multiple genes which behave like housekeeping genes.

3.1.5 Gene expression level in the heart and association to cardiac phenotypes

The availability of a measured ratio of the expression level of a gene in the heart vs. non-cardiac tissues provided means to formally test the hypothesis that the mutation in a gene with a typical high expression level in the heart in comparison with non-cardiac tissues will result in a pathologic heart-related phenotype. For this reason, we chose the set of potassium channel encoding genes, due to their relevance for normal heart functioning. For each gene named by a

KCN root, the certain or suspected association to a cardiac phenotype was recorded by manual searching of the OMIM database (Amberger et al., 2015). The significance of the association of a differential expression in the heart in comparison to non-cardiac tissues to the description or not of a cardiac phenotype when each gene is mutated was tested by Fisher exact test.

3.1.6 Gene selection for map validation

In order to obtain a sample experimental confirmation of the meta-analysis derived maps, for each tissue we selected a group of genes with these features: range of expression values covering the whole range of the expression magnitude order as calculated by TRAM; regular spacing of the expected expression values, i.e. each gene is expected to have a fold increase detectable through "Real Time" RT-PCR (at least 1 PCR cycle) in comparison with the subsequent gene with a lower expression in the group; the gene is known and characterized; when possible, the chosen gene is known to have a specific function in each tissue.

3.1.7 RT-PCR (Reverse transcriptase - polymerase chain reaction)

cDNA (complementary DNA) templates were obtained from reverse transcription of commercial human brain total RNA, human cerebellum total RNA, human cortex total RNA, human hippocampus total RNA and human heart total RNA (Clontech, Mountain View, CA). They were derived from the normal whole brain of an 18 year old Caucasian male, normal cerebellum pooled from 10 male/female Caucasians, age 22-68 years, normal cerebral cortex pooled from 5 male Asians, age 20-44 years, normal hippocampus of a 27 years old Asian male, normal whole heart pooled from 3 male Caucasians, aged between 30 and 39 years.

Reverse transcription conditions used were: 4 µg of total RNA (1 µg/µL), SuperScriptIII First-strand Synthesis Supermix (Invitrogen by Life Technologies, Grand Island, NY, USA) containing RT enzyme mix (includes SuperScript III RT 200 U/µL) 8 µL and RT reaction mix (includes oligo dT-20 2.5 µM, random hexamers 2.5 ng/µL, $MgCl_2$ 10 mM and dNTPs 10 mM) 40 µL. The RT-PCR reaction was performed in a final volume of 80 µL to have the same template for all the subsequent reactions.

The reaction consisted of three steps: an incubation of 10 min at 25°C, followed by an incubation at 50°C for 30 min and a final step of 5 min at 85°C. *E.coli* RNase H 4 µL (8 U) was then added to the reaction for 20 min at 37°C.

3.1.8 Primer design

Primers pairs were designed with "Amplify 3" software (Engels 1993) following standard criteria (Sharrocks, 1994). They are designed to specifically recognise expressed sequences (each primer being designed on a different exon) and to bind to regions common to all isoforms of the same gene because microarray probe sequences complement the known isoform sequences of the same gene. These constrains cause a variation in the amplicon lengths between 82 and 247 bp. Each primer is about 20-23 nt long, with an annealing temperature of 61° C.

3.1.9 PCR (polymerase chain reaction)

First, a qualitative analysis of reverse transcription products was performed using PCR and agarose gel electrophoresis. PCR experiments were performed in a 25 µL final volume, containing 2.5 µL of cDNA, 1 U Taq polymerase (TaKaRa, Shiga, Japan) with companion reagents (0.2 mM each dNTPs, 2 mM $MgCl_2$, 10× PCR buffer) and 0.2 µM of each primer. An initial denaturation step of 2 min at 94°C was followed by 25 cycles of 30 sec at 94°C, 30 sec at annealing temperature (Ta 61°C), 30 sec at 72°C, and a final extension of 7 min at 72°C. All RT-PCR products obtained were gel analyzed following a standard method (Davis et al., 1994).

3.1.10 Real-Time PCR profile and melting curve

Real-Time PCR assays were performed in triplicate, using the CFX96 instrument (Bio-Rad Laboratories, Hercules, CA). The reactions were performed in a total volume of 20 µL containing: 2.5 µL of cDNA (heart, brain, cerebral cortex, cerebellum final concentration: 0.78 ng/µL; hippocampus final concentration: 3.125 ng/µL); 10 µL of Sybr Select Master Mix 2× for CFX (Applied Biosystem, by Life Technologies) containing AmpliTaq® DNA Polymerase, UP (high purified), SYBR® GreenER™ dye and Heat-labile uracil-DNA glycosylase (UDG); 0.8 µL (0.3 µM) of both forward and reverse primer (MWG, Life Technologies); 5.9 µL of RNase-free water. Cycling parameters were: 2 min at 50°C (UDG activation), 2 min at 95°C (AmpliTaq Fast DNA Polymerase UP activation), 40 cycles of 15 sec at 95°C (denature) and of 1 min at 60°C (anneal and extend). A melting step needs to be performed to assay amplification specificity. This step consisted of an increase in temperature of 0.5°C/sec from 65°C to 95°C.

For each gene we used the primer pair that gave between 90-110 % efficiency. We used the ΔC_t (delta cycle threshold) method, a variation of the Livak method (Livak and Schmittgen, 2001), that uses the difference between reference and target gene C_t values for each sample to do

a relative quantification normalized to a reference gene (Observed Ratio(reference/target) = $2^{Ct(reference) - Ct(target))}$. For each gene group we set the gene with an intermediate expression value and a low standard deviation (SD, expressed as percentage of the mean value) in TRAM analysis as reference gene. The ratio among the transcriptome map expression values was calculated by dividing each expression value of the target gene by the expression value of the reference (expected ratio). Then we compared this value with the observed ratio and we examined the relationship between these two variables through bivariate statistical analysis using JMP 5.1 software (SA Institute, Campus Drive Cary, NC).

3.2 DS AMKL transcriptome

3.2.1 AMKL gene expression literature search

A systematic biomedical literature search was performed up to January 2013 in order to identify articles related to global gene expression profile experiments in AMKL patients (DS AMKL, non-DS AMKL and TMD conditions). A general search using the commonly used acronym "AMKL" retrieved 157 articles.

The MeSH term "Leukemia, Megakaryoblastic, Acute" was also used for a PubMed search in the expression: "Leukemia, Megakaryoblastic, Acute"[Mesh] AND ("Gene Expression Profiling"[MeSH] OR "Oligonucleotide Array Sequence Analysis"[Mesh] OR "Microarray Analysis"[Mesh] OR microarray* OR "Expression profile" OR SAGE).

3.2.2 AMKL gene expression database search

Gene Expression Omnibus (GEO) (Barrett and Edgar, 2006) functional genomics repository was searched for: (AMKL[All Fields] OR (AML[All Fields] AND M7[All Fields])) AND "Homo sapiens"[Organism]. A more general search using the expression "Down Syndrome"[MeSH] AND "Homo sapiens"[Organism] was also used.

ArrayExpress database (Sarkans et al., 2005) of functional genomics experiments was searched for the terms: "AMKL", "Megakaryoblastic", "AML M7".

In order to obtain gene expression profile datasets for normal human MK cells, in addition to the 9 used in the original description of the TRAM software (Lenzi et al., 2011), we searched GEO for the expression ("Megakaryocytes"[Mesh] OR Megakaryoblast*) AND "Homo sapiens"[ORGANISM]. The ArrayExpress database was searched for the expressions "Megakaryocyte", "Megakaryocytic", "Megakaryoblast", "MK".

The searches were performed up to January 2013.

3.2.3 AMKL gene expression dataset selection

The inclusion criteria of datasets in the analysis were: availability of the raw or pre-processed data; pediatric age of the subject from whom the sample was obtained; diagnosis of DS or non-DS AMKL or TMD.

Exclusion criteria were: exon arrays (hampering data elaboration by TRAM due to exceedingly high number of data rows) or platforms using probes split into several distinct arrays for each sample (hampering intra-sample normalization); lack of identifiers corresponding to those found in the GEO sample records (GSM) or Array Express sample records; platforms assaying an atypical number of genes (i.e. <5.000 or >60.000); cell line derived data; specific subtype of non-DS AMKL (e.g. t 1;22); trisomy 21 in non-DS AMKL samples.

Normal MK samples were considered for the analysis when fulfilling these criteria: late MK colonies (10-14 days) or MK sorted cells, obtained from peripheral blood (PB), bone marrow (BM) or cord blood (CB). MK cultured for less than 10 days or Colony Forming Unit-Megakaryocytic (CFU-MK) were excluded.

In order to obtain a quantitative transcriptome map, values from each dataset were linearized when provided as logarithms. In some cases we used raw files (e.g. File CEL) to be converted into pre-processed data, using the software "Alt Analyze" (Emig et al., 2010).

3.2.4 AMKL TRAM (Transcriptome Mapper) analysis

We created a directory (folder) for each condition, containing all the sample datasets related to the same source and selected for the study: DS AMKL (pool 'A'); non-DS AMKL (pool 'B'); TMD (pool 'C'); normal MK (pool 'D'); normal CB MK (pool 'E').

We ran the whole set of analyses permitted by TRAM (in both "Map" and "Cluster" mode, although we focused on the "Map" mode) using default parameters as described (Lenzi et al., 2011). We used an updated version of TRAM including enhanced resolution of gene identifiers and updated UniGene and Entrez Gene databases (TRAM 1.1, June 2013), in comparison with the original 2011 version (Lenzi et al., 2011). When the gene location cytoband was not available in the Gene database (Gene database (http://www.ncbi.nlm.nih.gov/gene)), it was manually derived from UCSC Genome Browser (University of California Santa Cruz (UCSC) Genome Browser. (http://genome-euro.ucsc.edu/cgi-bin/hgGateway)).

3.2.5 Other Analysis

FuncAssociate analysis (Berriz et al., 2003) was used to obtain Gene Ontology attributes in order to functionally characterize large sets of genes derived from the TRAM analysis.

3.3 HR-DSCR on human chromosome 21 and gene hunting

3.3.1 HR-DSCR computational biology analysis

It was possible to perform an analysis of computational biology in the highly restricted Down syndrome critical region (HR-DSCR) through the use of different programs which analyze the homology between nucleotide and amino acid sequences.

In order to verify the presence of coding sequences in the HR-DSCR, we analyzed successive sections of 10,000 bp within the region, including between the coordinates 37,929,229 to 37,963,130 (34 Kb), using the program BLASTX, without a search filter, which compares a nucleotide sequence to a database of proteins, translating it into all of the 6 possible reading frames. The sequence of interest was compared with the ESTs cataloged in the GenBank database (http://www.ncbi.nlm.nih.gov/genbank/) in order to verify the presence of transcripts not necessarily characterized as known genes. With the BLASTN program you can run an alignment of the sequence of interest (query) against the EST sequences. We entered as a query: NC_000021, with limits 37,929,229 - 37,963,130 placed in boxes labeled "from" "to", selecting as filters: "ESTs" and "Homo sapiens".

The results obtained by the search for homology (BLAST) were integrated using the program FGENESH (www.softberry.com/), which allows you to make an *ab initio* gene prediction in a fast and accurate way (Yu et al., 2002). This program, starting from a nucleotide sequence, detects the presence of the classical elements of gene functioning gene, such as: promoters, exons, start and termination codons of translation, etc. We started the *ab initio* research with FGENESH, inserting the genomic sequence corresponding to the region of interest of the human chromosome 21 (chr21). To validate the results of the prediction made with FGENESH you can use its variants, FGENESH +, FGENESHC and FEGENESH-2, which, through a homology search, comparing the prediction respectively with amino acid sequences, cDNA sequences and sequences orthologous notes, greatly improves the accuracy of the prediction of a gene sequence.

Finally, with the aim to check the presence of small regulatory sequences, miRNAs, within the HR-DSCR, we used the program miRBase (http://www.mirbase.org/). The region of 10,000

bp was divided into blocks of sequences of 1,000 bp and each block of sequences in FASTA format, was analyzed.

3.3.2 Putative miRNA1 Northern Blot analysis

Northern Blot allows the identification and quantification of specific RNA molecules due to hybridization with DNA probes (Alwine et al., 1977). We used this method to check for the putative precursor and mature miRNA1 forms. The method consists of six basic stages: preparation of a denaturant agarose gel for electrophoresis of RNA from different tissues; transfer of the RNA from the agarose gel to a membrane (filter) of nitrocellulose or nylon with surface charge; labelling with radioisotopes of a DNA probe complementary to the transcript of interest; hybridization of the probe to the filter; and exposure of the filter to an autoradiographic sheet resulting in the development of the same.

We used a commercial filter including poly-A RNA from twelve human tissues: brain, heart, skeletal muscle, colon, thymus, spleen, kidney, liver, intestine, placenta, lung and peripheral blood leukocytes. The filter named MTN2 ("multiple normal tissue", Clontech, MountainView, CA) consists of a nylon membrane with a charged surface.

To highlight the presence in the filter of the RNA target, it exploits the properties of nucleic acid molecules to pair with complementary sequences radioactively labelling the DNA sequences (radioactive probes) complementary to the target RNA. To label the probe, we added nucleotides labeled with phosphorus 32 (^{32}P) to the 3' end of the molecule. The reaction was performed with the enzyme terminal deoxynucleotidyl transferase (TdT) (Thermo Fischer Scientific, Waltham, Massachusetts, United States) in the presence of alpha ^{32}P dATP (3,000 Ci/mmol). The manufacturer's instructions suggest to perform the reaction in 50 µL of final volume with 10 pmol of the probe.

The TdT designed and labeled probes were probe miRNA1, which binds to all three forms of the putative miRNA1 (primary, precursor and mature) and pri-miRNA1 probe, complementary to the primary putative miRNA1 form (Table 2).

Probe	**Sequence**
miRNA1	GAGACAGAGTTTTGCTCTTGTTGCCC
pri-miRNA1	CCAGATTTGTTTTTATTTGATGTGTCTGGCCTCACTTGCTCAGCATGATGG

Table 2. Probes used to perform Northern Blot hybridization.

The labeled probe was purified using the "Illustrates Microspin G-25 Columns" kit (GE Healthcare, Little Chalfont, UK) according to the manufacturer's instructions. These columns allow to discard the nucleotides that were not incorporated into the probe during labelling. A qualitative analysis of the purification of the labeled probe was performed, checking the number of counts per minute (cpm) of the radioactive probe at this stage through a "beta-counter" measurement tool. The measurement was made on 1 µL of probe in 3 mL of scintillation liquid to determine the volume of the probe to be added in the hybridization solution for a total of 2 Mcpm/mL.

For the hybridization reaction we prepared 30 mL of hybridization solution containing: 1 M EDTA pH 8, 250 mM Na_2HPO_4 pH 7.2, 7% SDS. 20 mL of hybridization solution were put in a bottle. The filters were soaked in sterile distilled water and subsequently placed in the bottle.

A pre-hybridization step was performed for 30 minutes at 65°C. Subsequently, the bottle was emptied and was added to 10 mL of hybridization solution and the volume of labeled probe corresponding to 20 Mcpm. Hybridization was performed over night at 65°C.

After the hybridization reaction, the filter was washed with a solution containing a lower percentage of SDS (5%) to eliminate the excess of radioactive probe that did not bind to the filter. Finally the filters were exposed to an autoradiographic sheet, within a specific radiographic cassette, and stored at -80 ° C. The plates were developed after 24 hours.

On the autoradiographic sheet, signals of hybridization of the probe are visible in the form of bands, the height of which corresponds to the size of the target transcript. For the determination of the size of the transcript, we used the molecular markers indicated on the filter MTN2 (Figure 1) as reference. We used a sheet of graph paper in logarithmic scale with the length in millimeters (mm) of the distance between each molecular marker and the point corresponding to the loading well on the abscissa and the size in Kb of each molecular marker (Figure 1) on the ordinates. Starting from the axiom that for two points passes only one straight line, it was possible to draw a straight line that passes through the points designated by the molecular weight markers. Reporting the distance between the band visible on the autoradiographic sheet and the well on the abscissa, it is possible to derive the point of intersection with the straight line, which will allow to extrapolate the length in Kb of the transcript.

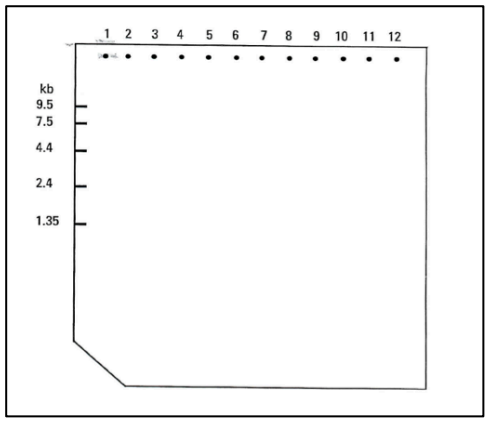


Figure 1 Diagram of commercial filter MTN2. The filter contains polyadenylated RNA from 12 human tissues: lane 1) brain, 2) heart, 3) skeletal muscle, 4) colon, 5) thymus, 6) spleen, 7) kidney, 8) liver, 9) intestine, 10) placenta, 11) lung, 12) peripheral blood leukocytes. The black dots indicate the loading well of RNA samples. To the left of the filter is shown the lengths in kb of molecular markers.

The hybridized filter was washed with a 0.1% solution of SDS brought to boiling to remove the probe remained tied.

Another method of detection of the hybridization signal on the filter is the exposure of the filter to a phosphorus screen (Storage Phosphor Screen, Kodak, Rochester, NY) and the acquisition of the image to PhophorImager (Storm 840, Molecular Dynamics, Amersham Pharmacia Biotech , Uppsala, Sweden). The activity of the radioisotope incorporated in the probe impresses the screen of phosphorus, which has a sensitivity of 10 to 100 times greater than the autoradiographic sheet. A few hours of exposure are needed to detect the hybridization signal impressed on the screen of phosphorus. The PhosphorImager is equipped with a helium or neon lamp that emits rays which excite the most common radioisotopes ^{14}C, ^{3}H, ^{125}I, ^{32}P, ^{33}P, ^{35}S. Radioisotopes, when excited, emit a phosphorescence that is collected by an optical fiber and channeled into a photomultiplier tube. The resulting data consists of positional information which, scanned with a precision of 16 bits, leads to the formation of an image.

3.3.3 Genomic organization of *DSCR4* locus

The locus *DSCR4* is adjacent to the HR-DSCR. For this reason we felt it essential to perform a bioinformatic study of the organization of this locus. To get an overview of the locus organization we used UCSC Genome Browser, a database in which the reference sequences and assembly projects that constitute a rich gene bank are stored. You can access the database from the link: "https://genome.ucsc.edu/" and enter the reference chromosome using a universal nomenclature in which the chromosome is referred to by the initials "chr" followed by the number of chromosome and two points (eg. chr21:), along with the start and end coordinates of

the locus to be studied, using thousands separator commas (37,929,229 - 37,963,130). After entering the coordinates, you can observe the genomic organization of the locus: the exons are shown as rectangles, the introns are shown as arrows, and the EST sequences, detectable by selecting "Human EST that have been spliced", are linked to an access code. We used the BLAST program "2 seq", which allows the alignment of the two sequences to compare EST sequences (highlighted in previous research) with the genomic sequence of chromosome 21, including as a reference their access codes. To find ESTs also partially homologous to the reference EST sequence, we used the program BLASTN, by entering the EST identification code as a query and the terms "Expressed sequence tags (est)" and "Homo sapiens" as filters.

Chapter 4

Results

4.1 Systematic meta-analysis of gene expression profiles of normal human tissues

4.1.1 Gene expression database search and database building

The performed search, followed by checking for exclusion and inclusion criteria as described in the "Materials and Methods" section above, retrieved 60 samples from 15 microarray experiments on whole adult brain, 35 samples from 13 microarray experiments on whole foetal brain, 140 samples from 15 microarray experiments on whole cerebellum, 18 samples from 4 microarray experiments on whole cerebral cortex, 41 samples from 7 microarray experiments on the whole hippocampus, 32 samples from 11 microarray experiments on the whole heart, 622 samples from 12 microarray experiments on non-brain tissue pool, and 629 samples from 13 microarray experiments on non-cardiac tissue pool. Datasets search on different human tissues retrieved ten articles describing gene expression profile for 53 different tissues or organs. Sample identifiers and main sample features of brain, foetal brain, cerebellum and cerebral cortex are listed in Supplementary Table 1 and for human non-brain comparison pool in Supplementary Table 2, both available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/. Sample identifiers and main sample features of hippocampus are listed in Supporting Information Table S1, available at: http://apollo11.isto.unibo.it/suppl/2015_Caracausi/. Sample identifiers and main sample features of whole heart and pool of non-cardiac tissues are listed in the Table I and II of the Data Supplement, both available at: http://apollo11.isto.unibo.it/heart, in the event of acceptance of the manuscript they will be available at: http://apollo11.isto.unibo.it/suppl/).

Datasets were loaded into TRAM and analyzed obtaining eighteen transcriptome maps: adult brain (pool 'A'); adult brain (pool 'A') vs. pool of tissues minus brain (pool 'B'); male brain (pool 'A.1'); female brain (pool 'A.2'); male brain (pool 'A.1') vs. female brain (pool 'A.2'); foetal brain (pool 'C'); foetal brain (pool 'C') vs. adult brain (pool 'A'); cerebellum (pool 'D'); cerebellum (pool 'D') vs. adult brain (pool 'A'); cerebral cortex (pool 'E'); cerebral cortex (pool 'E') vs. adult brain (pool 'A'); hippocampus (pool 'F'); hippocampus (pool 'F') vs. whole brain (pool 'A'); male hippocampus (pool 'F.1'); female hippocampus (pool 'F.2'); male hippocampus (pool 'F.1') vs. female hippocampus (pool 'F.2'); heart (pool 'G'); heart (pool 'G') vs. pool of tissues minus heart (pool 'H'). All the folders corresponding to each single transcriptome map are listed in the Table 1.

4.1.2 Transcriptome differential maps

Each map provides: the total of data points analyzed for each tissue, i.e. gene expression values (expressed as percentage of the mean value) for all human mapped loci following intra- and inter-sample normalization (Lenzi et al., 2011); the number of loci for each tissue and for which the comparison between two conditions (different tissues) was possible, due to the presence of values for those loci in both sample pools considered; the number (at least three over/under-expressed genes) and the gene content of each genomic segment found to be statistically significantly over/under-expressed in the comparison between the two tissues. Each genomic segment was identified among the 12,373 segments generated using the default window of 500,000 bp with a sliding window of 250,000 bp and following removal of overlapping segments with similar gene content.

A segment or a gene was considered to be statistically significantly over/under-expressed for $q<0.05$, where q is the p-value obtained by the method of hypergeometric distribution (Lenzi et al., 2011) and corrected for multiple comparison.

4.1.3 Adult brain and Adult brain vs. Pool of tissues minus brain transcriptome map analysis

In the adult brain transcriptome map analysis 1,803,680 data points corresponding to 39,250 mapped loci (Supplementary Table 3, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/) were included. 22,699 data points of the map correspond to 565 chr21 mapped loci. Results obtained by analysis included 28 significantly over-expressed segments.

The genome segment that has the highest statistically significant expression value is on chromosome 12 (12q13.12) (Table 3a), including the over-expressed known genes *TUBA1A*, *TUBA1B* and *TUBA1C*, encoding respectively for brain-specific tubulin alpha 1a, alpha 1b and alpha 1c. There are no statistically significant under-expressed segments.

Table 3 List of the three most **a)** over- or **b)** under- expressed genes and genomic segments (all significantly, with q<0.05). In **a)** all genes and segments are sorted by decreasing expression value, in **b)** by increasing expression value. Data refer to the transcriptome map of whole adult brain; whole foetal brain; cerebellum and cerebral cortex.

a)

TISSUE **SEGMENT OR GENE**	**Whole adult brain**	**Whole foetal brain**	**Cerebellum**	**Cerebral Cortex**
Over-expressed segments				
I	**12q13.12** 566.00	**12q13.12** 981.68	11q13.1 492.10	**12q13.12** 1051.22
	TUBA1B *TUBA1A* *TUBA1C*	*TUBA1B* *TUBA1A* *TUBA1C*	Hs.736281 Hs.593027 *MALAT1* Hs.712678 Hs.732685	*TUBA1B* *TUBA1A* *TUBA1C*
II	Xp11.23[a] 561.23	4q21.3[a] 702.21	6p21.3 490.85	12q12 785.70
	BEX1 *BEX4* *TCEAL5* *BEX2* *TCEAL7* *NGFRAP1*	*RPS29* *RPL36AL* *KLHDC2*	*GRM4* Hs.592692 *RPS10* *PACSIN1*	*ARF3* *DDN* *TUBA1B* *TUBA1A* *TUBA1C*
III	**6p21.3** 524.86	**6p21.3** 658.07	11q13.1[a] 487.38	11q13.1 738.20
	Hs.743967 Hs.592692 Hs.597332 *RPS10* *PACSIN1*	*C6orf1* Hs.743967 *RPS10*	*MALAT1* Hs.712678 Hs.732685 *CFL1*	Hs.736281 Hs.593027 *MALAT1* Hs.712678 Hs.732685
Over-expressed genes				
I	***BCYRN1*** 7,875.54	*TUBA1B* 6,954.44	***BCYRN1*** 12,372.42	Hs.732685 12,387.16
II	***TUBA1B*** 4,876.25	*TUBA1A* 6,810.31	Hs.732685 8,742.11	***TUBA1B*** 6,343.86
III	Hs.732685 4,680.24	*EEF1A1* 6,336.99	*CALM2* 4,704.56	*TUBA1C* 6,093.47
Over-expressed chr21 genes				
I	*SOD1* 1,727.39	***DNAJC28*** 2,817.20	***DNAJC28*** 1,447.08	*OLIG1* 726.82
II	*DNAJC28* 878.57	***SOD1*** 1,657.36	***SOD1*** 1,432.22	*PCP4* 700.08
III	*APP* 791.87	*ATP5O* 1,074.36	*TIAM1* 1,229.18	*SOD1* 629.70

b)

TISSUE SEGMENT OR GENE	Whole adult brain	Whole foetal brain	Cerebellum	Cerebral Cortex
Under-expressed segments				
I			13q21.31[a] 3.19 Hs.375745 Hs.735749 Hs.551057	
Under-expressed genes				
I	*TRG* 2.93	Hs.439634 0.17	Hs.707129 0.99	*ZNF852* 3.59
II	Hs.737002 3.00	Hs.674562 0.73	Hs.680393 1.01	Hs.28723 3.71
III	Hs.729885 3.73	Hs.599650 1.01	*OR2T29* 1.1	Hs.441636 3.74
Under-expressed chr21 genes				
I	Hs.542623 5.08	Hs.729539 2.31	Hs.561029 1.82	Hs.50927 5.48
II	Hs.561029 5.13	***LOC339622*** 3.10	Hs.677645 1.89	***LOC339622*** 6.29
III	Hs.542565 5.18	Hs.666775 3.16	Hs.580903 1.93	Hs.290805 6.44

Analysis was performed using default parameters (see "Methods" section). Under each segments the expression value and the genes included in the segment are indicated, under each gene the expression value is indicated. Some segments are not shown for simplicity because they are over-lapping with those highlighted in one of the listed regions. The listed gene for each segments are only the significantly over-/under-expressed genes. The complete results for these models are available as online supplementary material. In bold are marked the results overlapping among the transcriptome maps. In the 'Map' mode, TRAM displays UniGene EST clusters (with the prefix "Hs." in the case of H. sapiens) only if they have an expression value. [a] Cytoband was derived from the UCSC Genome Browser.

At single gene level, *BCYRN1* (chr2), encoding for a brain cytoplasmic RNA 1, has the highest expression value (7,875.54) (Table 3a); *TRG* (chr7), encoding for T cell receptor gamma, has the lowest expression value (2.93) (Table 3b). Among the chr21 genes *SOD1*, encoding for superoxide dismutase 1 soluble, has the highest expression value (1,727.39), followed by *DNAJC28* (878.57), encoding for DnaJ (Hsp40) homolog, subfamily C, member 28, and *APP* (791.87), encoding for amyloid beta (A4) precursor protein (Table 3a).

In the analysis of the Adult brain vs. Pool of tissues minus brain TRAM map, regional differential expression of pool 'A' (60 total brain samples) versus pool 'B' (622 pool of tissues minus brain samples and 34,985 mapped loci listed in the Supplementary Table 4, available at:

http://apollo11.isto.unibo.it/suppl/2014_Caracausi/) was investigated. Results included 23 significantly over- (n=7) or under-expressed (n=16) segments.

The genome segment that has the highest statistically significant expression value is on chromosome 5 (5q34) (Table 4a), including the over-expressed known genes *GABRA6*, *GABRA1* and *GABRG2*, encoding respectively for gamma-aminobutyric acid (GABA) A receptor, subunit alpha 6, gamma-aminobutyric acid (GABA) A receptor, subunit alpha 1 and gamma-aminobutyric acid (GABA) A receptor, gamma 2. The genome segment that has the lowest statistically significant expression value is on chromosome 2 (2p12) (Table 4b), including the under-expressed known genes *REG1B*, *REG1A* and *REG3A*, encoding respectively for regenerating islet-derived 1 beta, regenerating islet-derived 1 alpha and regenerating islet-derived 3 alpha.

At single gene level, an increase of more than 10 times was observed in all the first 125 loci (Supplementary Table 5, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/). In particular, a fold increase of 79.69 was observed for the known gene *ANKRD30B* (Table 4a), encoding for ankyrin repeat domain 30B. We also observed that in this range of expression ratio, five chr21 genes are included: *DNAJC28*, *LINC00320*, *LINC00323*, *OLIG1* and *OLIG2* (Supplementary Table 5, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/). In particular, *DNAJC28* had a 65.06-fold increase with respect to all tissues, followed by *LINC00320*, encoding for a long intergenic non-protein coding RNA 320, that had a 39.82-fold increase (Table 4a).

Among the genes with the lowest 'A'/'B' expression ratio, a fold decrease of 100 was observed for the EST clusters Hs.633942, Hs.554169, Hs.727036, Hs.720702 and for the known genes *CSTA*, *KRT13* and *MSMB*, encoding for cystatin A (stefin A), keratin 13 and beta-microseminoprotein, respectively (Supplementary table 5, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/).

When the 24 spinal cord samples were removed from the non-brain sample pool, using the "Exclude" sample function provided by the TRAM graphical interface in order to test if the removal of a central nervous system organ from the human tissue set could alter the transcriptome comparison of the brain with non-brain tissue pool, substantially analogous results were obtained (data not shown).

Table 4 List of the three most **a)** over- or **b)** under- expressed genes and segments (all significantly, with q<0.05). In **a)** all genes and segments are sorted by decreasing gene expression ratio, in **b)** by increasing gene expression ratio. Data refer to the differential transcriptome map: whole adult brain vs. pool of non-brain tissue; whole foetal brain vs. whole adult brain; cerebellum vs. whole adult brain and cerebral cortex vs. whole adult brain.

a)

TISSUE **SEGMENT OR GENE**	**Whole adult brain vs. Pool of tissue**	**Whole foetal brain vs. Whole adult brain**	**Cerebellum vs. whole adult brain**	**Cerebral Cortex vs. whole adult brain**
Over-expressed segments				
I	5q34 11.35	8q13.1[a] 7.51	15q22.2 5.23	1q23.1 17.55
	GABRA6 GABRA1 GABRG2	*LOC100130155* Hs.388788 *BHLHE22*	*RORA* Hs.660127 Hs.655820	*OR6Y1 OR6K2 OR6N1 PYHIN1*
II	13q21.32 9.25	2q33 6.99	12q22[a] 4.05	11q12.1 7.8
	PCDH9 Hs.656886 Hs.676018	*SATB2* Hs.151184 *SATB2-AS1*	Hs.585087 *BTG1* Hs.434392	*OR5F1 OR5T2 OR8U1*
III	3q24 6.86	10p14 [a] 4.06	1p31.3 3.87	11q11 7.33
	ZIC4 Hs.720460 *ZIC1*	Hs.734120 Hs.676705 Hs.735675 Hs.655681	*INADL* Hs.737385 Hs.673484	*OR5L2 OR5F1 OR5T2*
Over-expressed genes				
I	*ANKRD30B* 79.69	*TMSB15A* 63.15	*CBLN3* 81.34	*ZNF790* 225.06
II	*DNAJC28* 65.06	*DCX* 42.79	Hs.665664 17.65	Hs.197693 110.78
III	*CDR1* 57.38	Hs.712990 40.03	Hs.12316 16.97	Hs.594912 86.53
Over-expressed chr21 genes				
I	*DNAJC28* 65.06	*CXADR* 7.40	Hs.657183 6.83	*KRTAP13-2* 32.77
II	*LINC00320* 39.82	*LRRC3DN* 4.86	*ADAMTS5* 6.36	*KRTAP15-1* 7.10
III	*OLIG1* 11.12	Hs.675532 4.24	*TIAM1* 5.29	Hs.657999 6.49

b)

TISSUE SEGMENT OR GENE	Whole adult brain vs. Pool of tissue	Whole foetal brain vs. Whole adult brain	Cerebellum vs. whole adult brain	Cerebral Cortex vs. whole adult brain
Under-expressed segments				
I	2p12 0.16	2q31.1 0.38		
	REG1B *REG1A* *REG3A*	*LRP2* *BBS5* *KLHL41* Hs.593163		
II	12q21.3-q22 0.17	5q31 0.41		
	LUM Hs.539252 *DCN*	Hs.658232 *NR3C1* Hs.703520		
III	18q12.1 0.19	7q22.2 0.54		
	DSC3 *DSC2* *DSC1*	Hs.718842 *LOC100216546* Hs.656426 Hs.657627		
Under-expressed genes				
I	Hs.633942 0.01	*TNNC1* 0.01	*NRGN* 0.01	Hs.683165 0.01
II	Hs.554169 0.01	Hs.80714 0.01	*LCE5A* 0.01	*CKM* 0.01
III	*CSTA* 0.01	*GGT6* 0.01	*CKM* 0.01	*DNAJC28* 0.02
Under-expressed chr21 genes				
I	*COL6A2* 0.09	***LINC00320*** 0.03	***LINC00320*** 0.04	*DNAJC28* 0.02
II	Hs.663673 0.11	*S100B* 0.05	*KRTAP10-10* 0.11	*LINC00320* 0.10
III	*FAM3B* 0.14	*C21orf91* 0.08	***LINC00323*** 0.15	***LINC00323*** 0.11

Analysis was performed using default parameters (see "Methods" section). Under each segments the expression ratio and the genes included in the segment are indicated, under each gene the expression ratio is indicated. Some segments are not shown for simplicity because they are over-lapping with those highlighted in one of the listed regions. The listed gene for each segments are only the significantly over-/under-expressed genes. The complete results for these models are available as online supplementary material. In bold are merked the results overlapping among the transcriptome maps. In the 'Map' mode, TRAM displays UniGene EST clusters (with the prefix "Hs." in the case of H. sapiens) only if they have an expression value. [a] Cytoband was derived from the UCSC Genome Browser.

4.1.4 Male brain vs. Female brain transcriptome map analysis

The adult brain samples for which the sex of the sample donor was available were grouped into two additional datasets: pool 'A.1' including 13 male samples and pool 'A.2' including 5 female samples and the corresponding transcriptome maps were generated and compared.

The gene expression value for each of the 27,437 loci of the male brain transcriptome map (Supplementary Table 6, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/) and for each of the 25,954 loci of the female brain transcriptome map (Supplementary Table 7, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/) are available. The differential transcriptome map produced a gene expression ratio for each of the 25,954 shared loci (Supplementary Table 8, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/). Results included 6 significantly over-expressed segments. The genome segment that has the highest statistically significant expression value is on chromosome Y, including the known genes *TTTY15*, *USP9Y* and *DDX3Y* encoding respectively for testis-specific transcript, Y-linked 15 (non-protein coding), ubiquitin specific peptidase 9, Y-linked and *DEAD* (Asp-Glu-Ala-Asp) box helicase 3, Y-linked (supplementary full results for transcriptome maps are available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/)

At single gene level, a more than ten-fold increase was observed in all the first 84 loci (Supplementary Table 8, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/). In particular, a more than 100-fold increase was observed for the uncharacterized locus *LOC613126* (378.56) and the known genes *LBX1* (366.49), *DRD4* (332.51) and *CDC42EP5* (115.26), encoding respectively for ladybird homeobox 1, D4 subtype of the dopamine receptor and for the *CDC42* effector protein (Rho GTPase binding) 5 (Table 5).

Among the genes with the lowest 'A.1'/'A.2' expression ratio, a 25-fold decrease was observed for the known gene *XIST*, encoding for the X inactive specific transcript (non-protein coding) (Table 5).

Table 5 List of the ten most over- and under-expressed genes in male brain (pool 'A.1') vs. female brain (pool 'A.2').

Gene name	Value 'A.1'	Value 'A.2'	Ratio 'A.1'/'A.2'	Chr	Data points 'A.1'	Data points 'A.2'	SD as % of expression 'A.1'	SD as % of expression 'A.2'
Over-expressed genes								
LOC613126	2,019.28	5.33	378.56	chr7	6	3	110.66	71.16
LBX1	529.57	1.44	366.49	chr10	16	5	214.23	75.11
DRD4	1,268.15	3.81	332.51	chr11	16	5	213.44	125.06
CDC42EP5	1,433.81	12.44	115.26	chr19	6	3	109.49	88.48
LOC100133315	578.64	7.97	72.59	chr11	9	3	148.30	106.70
EDARADD	126.93	1.89	67.15	chr1	6	3	103.76	35.84
HIST3H3	182.62	2.73	66.94	chr1	13	5	182.15	70.33
HES7	1,244.07	18.79	66.22	chr17	6	3	110.37	48.39
PPIAL4A	279.32	4.75	58.77	chr1	25	5	110.99	120.25
FTHL17	154.68	2.78	55.58	chrX	6	3	117.92	4.55
Under-expressed genes								
CCL4	27.16	201.40	0.13	chr17	13	5	54.58	172.86
IL1B	27.82	208.20	0.13	chr2	50	10	92.63	112.94
Hs.434622	0.90	6.78	0.13	chr1	3	3	50.62	63.74
Hs.611927	2.23	19.15	0.12	chr6	3	3	70.55	85.33
Hs.25345	0.38	3.42	0.11	chr6	3	3	19.05	44.69
NPAS4	8.37	80.07	0.10	chr11	9	6	144.65	158.20
LOC400768	0.41	4.24	0.10	chr1	3	3	8.38	57.81
CCL3	19.43	201.78	0.10	chr17	13	5	136.51	140.44
Hs.710548	0.75	7.94	0.09	chr17	3	3	27.09	63.80
XIST	7.97	203.97	0.04	chrX	44	31	78.75	131.58

Value: mean gene expression value normalized across all the pool samples; Data points: number of samples with an expression value for the locus; SD: standard deviation for the expression value expressed as a percentage of the mean. All these gene result statistically significant in the transcriptome map with a segment window of 12,500 bp. Full results available as supplementary material (see text).

4.1.5 Foetal brain and Foetal brain vs. Adult brain transcriptome map analysis

In the foetal brain transcriptome map analysis, 855,662 data points corresponding to 38,483 mapped loci were included (Supplementary Table 9, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/). 11,583 data points of the map correspond to 559 chr21 mapped loci. Results obtained by analysis included 38 significantly over-expressed segments.

The genome segment that has the highest statistically significant expression value is on chromosome 12 (Table 3a), including the over-expressed known genes *TUBA1A*, *TUBA1B* and *TUBA1C*. There are no statistically significant under-expressed segments.

At single gene level, *TUBA1B* (chr12), has the highest expression value (6,954.45) (Table 3a); the EST cluster Hs.439634 (chr14) has the lowest expression value (0.17) (Table 3b). Among the genes of chromosome 21 (chr21), *DNAJC28*, has the highest expression value (2,817.20); followed by *SOD1* (1,657.36) and *ATP5O* (1,074.36), encoding for ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit (Table 3a).

In the analysis of the Foetal brain vs. Adult brain TRAM map, regional differential expression of pool 'C' (35 foetal brain samples) versus pool 'A' (60 adult brain samples) was investigated. Results included 22 significantly over- (n=19) or under-expressed (n=3) segments.

The genome segment that has the highest statistically significant expression value is on chromosome 8 (Table 4a), including the over-expressed known gene *BHLHE22*, encoding for basic helix-loop-helix family, member e22, the uncharacterized locus *LOC100130155* and the EST cluster Hs.388788. The genome segment that has the lowest statistically significant expression value is on chromosome 2 (2q31.1) (Table 4b), including the under-expressed known genes *LRP2*, *BBS5*, *KLHL41*, encoding for low density lipoprotein receptor-related protein 2, Bardet-Biedl syndrome 5, and kelch-like family member 41 respectively, and the EST cluster Hs.593163.

At single gene level, an increase of more than 10 times was observed in all the first 44 loci (Supplementary Table 10, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/), including *TMSB15A*, encoding for thymosin beta 15a, and *DCX*, encoding for doublecortin, which have a fold increase of 63.15 and 42.79 (Table 4a), respectively.

All the chr21 loci have an increase of less than 10 times, and an increase between 2 and 10 times was observed for 28 loci (Supplementary Table 10, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/). In particular, *CXADR*, encoding for coxsackie virus and adenovirus receptor, has a 7.4-fold increase compared to adult brain (Table 4a). *LINC00320* has the lowest 'C'/'A' expression ratio (0.03) preceded by the known gene *S100B* (0.05), encoding for S100 calcium binding protein B (Table 4b). Among the genes with the lowest 'C'/'A' expression ratio a fold decrease of 100 was observed for the EST clusters Hs.80714, and for the known genes *LCE5A*, *ANKK1*, *GGT6* and *TNNC1*, encoding respectively for late cornified envelope 5A, ankyrin repeat and kinase domain containing 1, gamma-glutamyltransferase 6 and troponin C type 1 (slow) (Supplementary Table 10, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/).

4.1.6 Cerebellum and Cerebellum vs. Adult brain transcriptome map analysis

In the analysis of the cerebellum TRAM map, 3,862,253 data points corresponding to 38,163 mapped loci were included (Supplementary Table 11, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/). 49,004 data points of the map correspond to 554 chr21 mapped loci. Results obtained by analysis included 28 significantly over-expressed segments and 1 significantly under-expressed segment.

The genome segment that has the highest statistically significant expression value is on chromosome 11 (11q13.1) (Table 3a), including four over-expressed UniGene EST clusters (Hs.736281, Hs.593027, Hs.712678, Hs.732685) and the known gene *MALAT1*, encoding for metastasis-associated lung adenocarcinoma transcript 1 (non-protein coding). The genome segment that has the lowest statistically significant expression value is on chromosome 13 (Table 3b), including the under-expressed EST clusters Hs.375745, Hs.735749 and Hs.551057.

At single gene level, *BCYRN1* has the highest expression value (12,372.42) (Table 3a); the UniGene EST cluster Hs.707129 has the lowest expression value (0.99) (Table 3b). Among the chr21 genes, *DNAJC28* has the highest expression value (1,447.08), followed by *SOD1* (1,432.22) (Table 3a).

In the analysis of the Cerebellum vs. Adult brain TRAM map, regional differential expression of pool 'D' (140 cerebellum samples) versus pool 'A' (60 total brain samples) was investigated. Results included 42 significantly over-expressed segments.

The genome segment that has the highest statistically significant expression value is on chromosome 15 (15q22.2) (Table 4a), including the over-expressed known gene *RORA*, encoding for RAR-related orphan receptor A, and the two UniGene EST clusters Hs.660127 and Hs.655820. There are no statistically significant under-expressed segments.

At single gene level, an increase of more than 10 times was observed in all the first 23 loci, including the gene *FSTL5*, encoding for follistatin-like 5 (Supplementary Table 12, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/). In particular, a fold increase of 81.34 was observed for the known gene *CBLN3*, encoding for cerebellin 3 precursor (Table 4a). Among the genes with the lowest 'D'/'A' expression ratio a fold decrease of 100 was observed for the known gene *NRGN*, encoding for neurogranin (protein kinase C substrate, RC3) (Table 4b). All the chr21 loci have an increase of less than 10 times, and an increase between 2 and 10 times was observed for 40 loci in particular. The *DNAJC28* and *LINC00320* genes mentioned above have a 1.65-fold change and a 0.04-fold change respectively compared to total brain (Supplementary Table 12, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/).

4.1.7 Cerebral cortex and Cerebral cortex vs. Adult brain transcriptome map analysis

In the analysis of the cerebral cortex TRAM map 780,661 data points corresponding to 27,504 mapped loci were included (Supplementary Table 13, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/). 9,846 data points of the map correspond to 336 chr21 loci. Results obtained by analysis included 21 significantly over-expressed segments.

The segment that has the highest statistically significant expression value is on chromosome 12 (12q13.12) (Table 3a), including the over-expressed known genes *TUBA1A*, *TUBA1B* and *TUBA1C*. There are no statistically significant under-expressed segments.

At single gene level, the EST cluster Hs.732685 has the highest expression value (12,387.16), followed by *TUBA1B*, which is the first known gene with the highest expression value (6,343.86) (Table 3a). The known gene *ZNF852* has the lowest expression value (3.59) (Table 3b). Among the chr21 genes, *OLIG1*, encoding for oligodendrocyte transcription factor 1, has the highest expression value (726.82), followed by *PCP4*, encoding for Purkinje cell protein 4 (700.08), and *SOD1* (629.7) (Table 3a).

In the analysis of the cerebral cortex vs. adult brain TRAM map, regional differential expression of pool 'E' (18 cortex samples) versus pool 'A' (60 total brain samples) was investigated. Results included 10 significantly over-expressed segments.

The segment that has the highest statistically significant expression value is on chromosome 1 (1q23.1), including the over-expressed known genes *OR6Y1*, *OR6K2* and *OR6N1* encoding olfactory receptors, and *PYHIN1*, encoding pyrin and HIN domain family, member 1 (Table 4a). There are no statistically significant under-expressed segments.

At single gene level, an increase higher than 10 times was observed in all the first 178 loci (Supplementary Table 14, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/). In particular, an increase of 255.06 times was observed for the gene *ZNF790*, encoding for zinc finger protein 790 (Table 4a). Among the genes with the lowest 'E'/'A' expression ratio, a fold decrease of 100 was observed for the EST cluster Hs.683165 and for the known gene *CKM*, encoding for creatine kinase, muscle (Table 4b). We also observed that only one chr21 gene, *KRTAP13-2*, encoding for keratin associated protein 13-2, has an increase of 32.77 times, while the other chr21 loci have an increase of less than 7.1 times (Table 4a). The *DNAJC28* gene has the lowest 'E'/'A' expression ratio (0.02), preceded by *LINC00320* (0.1), both previously mentioned (Table 4b). Furthermore, we noted that *SOD1*, despite being among the genes of chr21 with a high expression value, has a low expression ratio compared to total brain (0.36) (Supplementary Table 14, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/).

4.1.8 Hippocampus and Hippocampus vs. Whole brain transcriptome map analysis

In the hippocampus transcriptome map analysis, 1,339,820 data points corresponding to 30,739 mapped loci were included (Supporting Information Table S2, available at: http://apollo11.isto.unibo.it/suppl/2015_Caracausi/, where the Pool 'F' here is named 'A'). 16,996

data points of the map corresponded to 403 chr21 mapped loci. Results obtained by analysis included 18 significantly over-expressed segments.

The genome segment that has the highest statistically significant expression value is on chromosome 12 (12q13.12) (Table 6), including the over-expressed known genes *TUBA1B*, *TUBA1A* and *TUBA1C*, encoding respectively for brain-specific tubulin alpha 1a, alpha 1b and alpha 1c. There are no statistically significant under-expressed segments.

Table 6 The first five genomic segments significantly over-/under-expressed in the hippocampus (pool 'F') transcriptome map.

Hippocampus transcriptome map					
Chr[a] and Location	**Segment Start[b]**	**Segment End[b]**	**Value 'F'**	**q-value**	**Genes in the segment[c]**
Over-expressed segments					
chr12 12q13.12	49,500,001	50,000,000	804.67	0.0098	***TUBA1B*, *TUBA1A*, *TUBA1C***
chr8 8p21.2	26,000,001	26,500,000	657.94	0.00088	***BNIP3L*, *PNMA2*, *DPYSL2***
chrX Xq22.1-q22.2[d]	102,250,001	102,750,000	653.12	0.0004	***BEX1*, *BEX4*, *BEX2*, *NGFRAP1***
chr12 12q12	49,250,001	49,750,000	614.69	0.00087	***ARF3*, *DDN*, *TUBA1B*, *TUBA1A*, *TUBA1C***
chr11 11q13.1	65,250,001	65,750,000	492.32	0.011	***MALAT1*, Hs.712678, Hs.732685, *CFL1***

Analysis was performed using default parameters (see "Materials and Methods" section). Segments are sorted by decreasing expression value in the hippocampus transcriptome map. In the 'Map' mode, TRAM displays UniGene EST clusters (with the prefix "Hs." in the case of *H. sapiens*) only if they have an expression value. For simplicity, some segments are not shown because they overlap with those highlighted in one of the listed regions. The complete results for these models are available as online supplementary material (see text). The segment location cytoband was derived from that of the first mapped gene within the segment. [a] Chromosome. [b] Segment Start/End: chromosomal coordinates for each segment. [c] Significantly over-/under-expressed genes as marked in the TRAM results. [d] Cytoband not available in Gene was derived from the UCSC Genome Browser (http://genome-euro.ucsc.edu/cgi-bin/hgGateway).

At single gene level, the EST cluster Hs.732685 (mapping within the known locus *MALAT1*, metastasis associated lung adenocarcinoma transcript 1 (non-protein coding), as shown by BLASTN analysis) (chr11) has the highest expression value (8,566.51), followed by the known genes *CALM2* (5,801.15), *TUBA1C* (4,677.64) and *TUBA1B* (4,488.91), encoding respectively for calmodulin 2 (phosphorylase kinase, delta), tubulin, alpha 1c and tubulin, alpha 1b (Table 7). The EST cluster Hs.655406 (mapping within the known locus *PCGF3*, polycomb group ring finger 3) has the lowest expression value (6.74). Among the genes of chromome 21, the long intergenic non-protein coding RNA, *LINC00114*, has the highest expression value (2,304.90), followed by *SOD1* (1,647.81) and *OLIG1* (872.42) (Table 7).

Table 7 List of the five most over- and under-expressed genes (all significantly, with q<0.05, except genes with [b] note) in the hippocampus (pool 'F') transcriptome map.

Gene name	Value 'F'	Location	Data points 'F'	SD as % of expression 'F'
Over-expressed genes				
Hs.732685[a]	8,566.51	11q13.1	24	35.69
CALM2	5,801.15	2p21	32	48.89
TUBA1C	4,677.64	12q13.12	40	37.64
TUBA1B	4,488.91	12q13.12	77	44.32
RPL41[b]	4,205.15	12q13	26	45.08
Over-expressed chr21 genes				
LINC00114	2,304.90	21q22.2	11	51.35
SOD1	1,647.81	21q22.11	41	28.92
OLIG1	872.42	21q22.11	25	56.16
TTC3	680.61	21q22.2	205	78.35
APP	566.72	21q21.3	106	76.46
Under-expressed genes				
Hs.655406[b]	6.74	4p16.3	5	31.25
Hs.542157[c]	7.54	2p24.1	5	40.10
LOC100505878	8.90	5q14.3	17	37.28
Hs.559317[c]	9.14	2q24.1	24	35.20
CSTF3-AS1	9.26	11p13	24	36.29
Under-expressed chr21 genes				
LOC339622	11.39	21q21.1-q21.2	24	34.27
Hs.434909[b]	12.65	21q22.12	17	26.44
Hs.290805	12.87	21q21.1	24	49.36
LINC00314	13.05	21q21.3	25	50.21
Hs.723551[b]	13.15	21q22.11	24	40.58

Value: mean gene expression value normalized across all the pool samples; Data poins: number of spots associated to an expression value for the locus; SD: standard deviation for the expression value expressed as a percentage of the mean. Over-expressed genes are sorted by decreasing gene expression value, under-expressed genes are sorted by increasing gene expression value. Full results are available as supplementary material (see text). [a] According to the criteria detailed in the Materials and Methods section, the segment window (12,500 bp) contains more than one gene, but the significance is assumed to remain because the expression value of this over- or under-expressed gene prevails over the others. [b] This gene is one of the five most over-/under-expressed ones, but the value is not statistically significant. This is either because this gene was present with other genes even in the 12,500 bp segment, and its expression value does not prevail over the others, or because the segment containing the gene does not fulfill criteria for local over-/under-expression, or because its expression value is not associated to at least 5 data points.[c] Cytoband not available in Gene was derived from the UCSC Genome Browser (http://genome-euro.ucsc.edu/cgi-bin/hgGateway).

In the analysis of the hippocampus vs. whole brain transcriptome map, regional differential expression of pool 'F' (41 hippocampus samples) versus pool 'A' (60 brain samples) was investigated. There are no statistically significant over-/under-expressed segments.

At single gene level, an increase of more than 10 times was observed in all the first 10 loci (Supporting Information Table S3, available at: http://apollo11.isto.unibo.it/suppl/2015_Caracausi/, where the Pool 'F' here is named 'A' and the Pool 'A' here is named 'B'), including *LINC00114* and the known gene *SLC44A4*, encoding for solute carrier family 44, member 4, which have increases of 304.38 and 24.58 (Table 8), respectively. We also observed that in this range of expression ratio, two chr21 genes are included: *LINC00114* (cited above) and *AIRE*, encoding for an autoimmune regulator, which has a 14.82-fold increase.

Among the genes with the lowest 'F'/'A' expression ratio, a 50-fold decrease was observed for the uncharacterized locus, *LOC613126*, and for the known gene, *LCE5A*, encoding for late cornified envelope 5A (Table 8).

Among the chr21 genes, *DNAJC28*, encoding for DnaJ (Hsp40) homolog, subfamily C, member 28, has the lowest 'F'/'A' expression ratio (0.03), followed by two long non-coding RNAs, *LINC00323* (0.12) and *LINC00320* (0.16) (Table 8).

Table 8 List of the five most over- or under-expressed genes (all significantly, with q<0.05, except genes with [a] note) in the hippocampus (pool 'F') vs. whole brain (pool 'A') transcriptome map.

Gene name	Value 'F'	Value 'A'	Ratio 'F'/'A'	Location	Data points 'F'	Data points 'A'	SD as % of expression 'F'	SD as % of expression 'A'
Over-expressed genes								
LINC00114	2,304.90	7.57	304.38	21q22.2	11	14	51.35	31.53
SLC44A4	269.98	10.98	24.58	6p21.3	80	74	261.98	41.67
ACTA1	248.91	15.10	16.48	1q42.13	41	47	387.16	48.11
IL10	151.36	9.42	16.07	1q31-q32	47	84	188.63	44.91
ANKRD20A1	193.75	12.07	16.06	9q13	30	29	153.49	115.47
Over-expressed chr21 genes								
LINC00114	2,304.90	7.57	304.38	21q22.2	11	14	51.35	31.53
AIRE	351.73	23.74	14.82	21q22.3	52	62	202.59	66.93
Hs.721038[a]	102.77	18.96	5.42	21q22.3	24	18	55.02	66.11
Hs.385528[b]	82.10	17.36	4.73	21q22.2	17	18	46.17	38.78
Hs.661896[a]	106.91	25.30	4.23	21q21.1	19	36	37.47	56.48
Under-expressed genes								
LOC613126	14.89	651.08	0.02	7q21.2	17	21	22.79	250.64
LCE5A[a]	22.50	946.89	0.02	1q21.3	10	14	46.27	108.63
DNAJC28	30.45	878.57	0.03	21q22.11	39	57	180.59	267.82
GGT6	25.65	714.15	0.04	17p13.2	10	25	20.25	173.34
ILDR1	22.35	544.13	0.04	3q13.33	58	58	39.07	350.19
Under-expressed chr21 genes								
DNAJC28	30.45	878.57	0.03	21q22.11	39	57	180.59	267.82
LINC00323	15.39	133.49	0.12	21q22.2	26	26	48.79	167.83
LINC00320	59.71	379.16	0.16	21q21.1	49	43	70.95	139.33
IFNAR1	52.06	148.85	0.35	21q22.11	89	102	59.67	484.24
LOC101060027[b]	90.45	257.07	0.35	21q22.11	24	18	30.82	65.02

Value: mean gene expression value normalized across all the pool samples; Data points: number of spots associated to an expression value for the locus; SD: standard deviation for the expression value expressed as a percentage of the mean. Over-expressed genes are sorted by decreasing gene expression ratio, under-expressed genes are sorted by increasing gene expression ratio. Full results are available as supplementary material (see text). [a] This gene is one of the five most over-/under-expressed ones, but the value is not statistically significant. This is either because this gene was present with other genes even in the 12,500 bp segment, and its expression value does not prevail over the others, or because the segment containing the gene does not fulfil criteria for local over-/underexpression, or because its expression value is not associated to at least 5 data points. [b] Cytoband not available in Gene was derived from the UCSC Genome Browser (http://genome-euro.ucsc.edu/cgi-bin/hgGateway).

4.1.9 Male hippocampus vs. Female hippocampus transcriptome map analysis

The hippocampus samples for which the sex of the sample donor was available were grouped into two additional datasets: pool 'F.1' including 15 male samples and pool 'F.2'

including 14 female samples, and the corresponding transcriptome maps were generated and compared.

The gene expression value for each of the 26,045 loci of the male and female hippocampus transcriptome map (Supporting Information Table S4, available at: http://apollo11.isto.unibo.it/suppl/2015_Caracausi/, where the Pool 'F.1' here is named 'A.1' and the Pool 'F.2' here is named 'A.2') is available. The differential transcriptome map produced a gene expression ratio for each of the 26,045 shared loci (Supporting Information Table S4, available at: http://apollo11.isto.unibo.it/suppl/2015_Caracausi/).

Results included 3 significantly over-expressed and 3 significantly under-expressed segments. The genome segment that has the highest statistically significant expression value is on chromosome Y (Yq11.21-q11.221), including the known genes *TTTY15*, *USP9Y* and *DDX3Y* encoding respectively for testis-specific transcript, Y-linked 15 (non-protein coding), ubiquitin specific peptidase 9, Y-linked and *DEAD* (Asp-Glu-Ala-Asp) box helicase 3, Y-linked (Table 9).

The genome segment that has the lowest statistically significant expression value is on chromosome X (Xq13.2), including the EST clusters Hs.720466 and Hs.648496 and the known genes *XIST*, *JPX* and *FTX*, encoding respectively for the X inactive specific transcript, the JPX transcript (XIST activator) and the FTX transcript (XIST regulator) (Table 9) (supplementary full results for transcriptome maps are available at: http://apollo11.isto.unibo.it/suppl/2015_Caracausi/).

Table 9 Genomic segments significantly over-/under-expressed in male hippocampus (pool 'F.1') vs. female hippocampus (pool 'F.2').

Chr[a] and Location	Segment Start[b]	Segment End[b]	Value 'F.1'/'F.2'	q-value	Genes in the segment[c]
Over-expressed segments					
chrY Yq11.21-q11.221[d]	14,750,001	15,250,000	2.54	0.000061	***TTTY15*, *USP9Y*, *DDX3Y***
chrY Yq11.221	22,250,001	22,750,000	2.25	0.000023	***TTTY10*, *EIF1AY*, Hs.670544**
Under-expressed segments					
chrX Xq13.2	73,000,001	73,500,000	0.72	0.00000014	***XIST*, *JPX*, Hs.720466, Hs.648496, *FTX*, Hs.625698**
chr6 6q25.3	158,500,001	159,000,000	0.78	0.00029	**Hs.633361, *TULP4*, *TMEM181***

Analysis was performed using default parameters (see "Materials and Methods" section). Over-expressed segments are sorted by decreasing expression ratio, under-expressed segments are sorted by increasing expression ratio. In the 'Map' mode, TRAM displays UniGene EST clusters (with the prefix "Hs." in the case of *H. sapiens*) only if they have an expression value. For simplicity, some segments are not shown because they overlap with those highlighted in one of the listed regions. The complete results for these models are available as online supplementary material (see text). [a] Chromosome. [b] Segment Start/End: chromosomal coordinates for each segment. [c] Significantly over-/under-expressed genes as marked in the TRAM results. [d] Cytoband not available in Gene was derived from the UCSC Genome Browser (http://genome-euro.ucsc.edu/cgi-bin/hgGateway).

At single gene level, a more than ten-fold increase was observed in all the first 7 loci (Supporting Information Table S4, available at: http://apollo11.isto.unibo.it/suppl/2015_Caracausi/). In particular, a more than 20-fold increase was observed for the known genes *CKM* (22.09) and *ACTA1* (20.20), encoding respectively for creatine kinase, muscle, and for actin, alpha 1, skeletal muscle (Table 10).

Among the genes with the lowest 'F.1'/'F.2' expression ratio, an approximately 3-fold decrease was observed for example for the known genes *HBG2* (0.30) and *XIST* (0.36), encoding respectively for the hemoglobin, gamma G, and for the X inactive specific transcript (Table 10).

Table 10 List of the ten most over- or under-expressed genes (all significantly, with q<0.05 except the genes with [a] note) in male hippocampus (pool 'F.1') vs. female hippocampus (pool 'F.2').

Gene name	Value 'F.1'	Value 'F.2'	Ratio 'F.1'/'F.2'	Location	Data points 'F.1'	Data points 'F.2'	SD as % of expression 'F.1'	SD as % of expression 'F.2'
Over-expressed genes								
CKM	387.51	17.55	22.09	19q13.32	15	14	247.78	30.96
ACTA1	609.79	30.19	20.20	1q42.13	15	14	254.15	33.04
RPS4Y1	370.50	20.97	17.67	Yp11.3	10	12	29.33	49.16
MB	501.88	30.76	16.31	22q13.1	15	14	249.69	36.76
MYL2	264.97	20.56	12.89	12q24.11	15	14	237.04	61.73
MYH2	247.98	20.09	12.34	17p13.1	15	14	240.86	57.98
TNNC2	363.55	31.10	11.69	20q12-q13.11	15	14	240.49	47.18
TNNC1	109.26	12.16	8.99	3p21.1	15	14	219.80	34.82
MYL1	176.50	21.52	8.20	2q33-q34	18	16	252.73	58.72
MYLPF	216.10	32.82	6.58	16p11.2	15	14	218.34	76.14
Under-expressed genes								
HBG2	81.26	270.10	0.30	11p15.5	19	22	67.38	175.46
HBG1[a]	47.30	141.97	0.33	11p15.5	15	14	46.22	198.87
XIST	215.35	606.58	0.36	Xq13.2	101	106	440.74	144.55
RASEF	160.55	445.96	0.36	9q21.32	50	56	144.19	140.79
FOLR1	287.61	717.29	0.40	11q13.3-q14.1	28	28	172.21	139.81
Hs.594321[a]	278.32	688.73	0.40	4p16.3	7	10	41.83	48.67
ANKRD36B	173.30	416.30	0.42	2q11.2	7	10	41.79	59.77
Hs.655766	68.35	155.26	0.44	16p13.3	7	10	50.36	19.13
Hs.744460	106.08	230.08	0.46	9q31.2	12	12	107.01	85.27
LOC100996667[b]	58.82	122.18	0.48	9p12	7	10	50.50	49.53

Value: mean gene expression value normalized across all the pool samples; Data points: number of spots associated to an expression value for the locus; SD: standard deviation for the expression value expressed as a percentage of the mean. Over-expressed genes are sorted by decreasing gene expression ratio, under-expressed genes are sorted by increasing gene expression ratio. Full results are available as supplementary material (see text). [a] This gene is one of the five most over-/under-expressed ones, but the value is not statistically significant. This is either because this gene was present with other genes even in the 12,500 bp segment, and its expression value does not prevail over the others, or because the segment containing the gene does not fulfil criteria for local over-/under-expression, or because its expression value is not associated to at least 5 data points. [b] Cytoband not available in Gene was derived from the UCSC Genome Browser (http://genome-euro.ucsc.edu/cgi-bin/hgGateway).

Bivariate analysis of the statistical correlation between the gene expression values of the hippocampus transcriptome map in male and in female cells was performed (Figure 2). The

results showed a significant statistical correlation of data (r=0.98, p-value<0.0001) confirming the large overlapping of results between the two transcriptome maps, with the exception of single genes with a well-known sex-biased expression pattern.

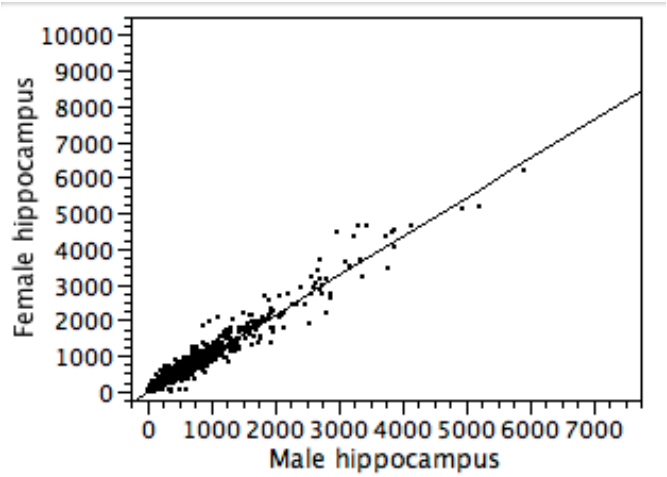


Figure 2 Bivariate analysis between the gene expression values of the hippocampus transcriptome map in male and in female cells (Supporting Information Table S4, available at: http://apollo11.isto.unibo.it/suppl/2015_Caracausi/). The fit line is shown; Pearson correlation coefficient is 0.98 and p-value<0.0001.

4.1.10 Heart and Heart vs. Pool of tissues minus heart transcriptome map analysis

In the heart transcriptome map analysis we check if the mean value provided by TRAM for the locus does not arise from a small number of extreme values included in the whole expression values attributed to the locus which may either be from biological variability or technical artefacts of hybridization due to the different probes in the different experimental platforms. In any case this possibility occurs in a smaller fraction of the loci (for example, only 3.8% of the loci has an SD>200 when expressed as % of the mean value, i.e. greater than twice the mean value).

Therefore, we decided to exclude from the results discussed in the manuscript those over- or under-expressed genes and segments containing genes whose expression values could be affected by a discrepancy between the values obtained with different probes assigned by the manufacturer to the same gene due to clear technical artefacts of hybridization.

In the analysis of the transcriptome map, 919,771 data points corresponding to 43,360 mapped loci were included (Table III in the Data Supplement, it is available at: http://apollo11.isto.unibo.it/heart, where the Pool 'G' here is named 'A'; in the event of acceptance of the manuscript it will be available at: http://apollo11.isto.unibo.it/suppl/). 11,474 data points of the map corresponded to 626 chr21 mapped loci. Results obtained by analysis included 36 significantly over-expressed segments.

The genome segment that has the highest statistically significant expression value is on chromosome 1 (1p36) (Table 11), including the over-expressed known genes *NPPA* and *NPPB* encoding respectively for the natriuretic peptide A and B, and *MFN2*, encoding for mitofusin 2.

There are no significantly under-expressed segments.

Table 11 The first five genomic segments significantly over-expressed in the whole heart (pool 'G') transcriptome map.

Whole heart transcriptome map					
Chr[a] and Location	**Segment Start[b]**	**Segment End[b]**	**Value 'G'**	**q-value**	**Genes in the segment[c]**
Over-expressed segments					
chr1 1p36	11,750,001	12,250,000	677.02	0.0085	***NPPA, NPPB, MFN2***
chr14 14q11.2	23,500,001	24,000,000	634.92	0.027	***PSMB5*, *MYH6*, *MYH7***
chr18 18p11.31	3,000,001	3,500,000	585.88	0.0022	***MYOM1*, *MYL12A*, Hs.569882, *MYL12B***
chr1 1q43	236,750,001	237,250,000	567.02	0.00046	***ACTN2*, Hs.601529, *MT1HL1*, *RYR2***
chr11 11q23.1	111,750,001	112,250,000	513.29	0.0016	***CRYAB*, *HSPB2*, *TIMM8B*, *SDHD*, Hs.356270**

Analysis was performed using default parameters (see "Results" section). Segments are sorted by decreasing expression value in the heart transcriptome map. In the 'Map' mode, TRAM displays UniGene EST clusters (with the prefix "Hs." in the case of *H. sapiens*) only if they have an expression value. For simplicity, some segments are not shown because they overlap with those highlighted in one of the listed regions. The complete results for these models are available as online supplementary material (see text). [a] Chromosome. The segment location cytoband was derived from that of the first mapped gene within the segment. [b] Segment Start/End: chromosomal coordinates for each segment. [c] Significantly over-/under-expressed genes as marked in the TRAM results.

At single gene level, the known gene *MYL2*, encoding for the myosin, light chain 2, regulatory, cardiac, slow protein, has the highest expression value (10,216.56), followed by the known genes *MYH7* (8,725.76) and *MB* (8,283.87), encoding respectively for myosin, heavy chain 7, cardiac muscle, beta protein and the myoglobin (Table 12). The EST cluster Hs.738508 has the lowest expression value (2.21).

Among the genes of chromosome 21, the known gene *ATP5J*, encoding for ATP synthase, H+ transporting, mitochondrial Fo complex, subunit F6, has the highest expression value (2,254.25), followed by *ATP5O* (2,205.50) and *SOD1* (1,212.50), encoding respectively for ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit and for superoxide dismutase 1, soluble (Table 12). The lowest expression value belongs to the EST cluster Hs.59471 (4.71) (Table 12).

Table 12 List of the five most over- and under-expressed genes (all significantly, with q<0.05, except genes with [a] note) in the whole heart (pool 'G') transcriptome map.

Whole heart transcriptome map				
Gene name	**Value 'G'**	**Location**	**Data points 'G'**	**SD as % of expression 'G'**
Over-expressed genes				
MYL2	10,216.56	12q24.11	33	41.46
MYH7	8,725.76	14q12	20	63.16
MB	8,283.87	22q13.1	39	59.23
TCAP[a]	6,352.89	17q12	20	112.52
NPPA[a]	6,332.35	1p36.21	18	77.72
Over-expressed chr21 genes				
ATP5J	2,254.25	21q21.1	36	84.88
ATP5O	2,205.50	21q22.11	34	72.59
SOD1[a]	1,212.50	21q22.11	58	52.38
Hs.129621[a]	1,128.12	21q22.3[b]	7	88.92
Hs.664516[a]	773.63	21q22.11[b]	7	258.34
Under-expressed genes				
Hs.738508[a]	2.21	8q24.12[b]	10	56.07
C3orf80	2.57	3q25.33	8	39.74
Hs.156135	2.58	1q32.1[b]	8	58.55
Hs.644952[a]	2.72	6q23.3[b]	5	36.55
Hs.637980[a]	2.82	11q13.4[b]	10	48.33
Under-expressed chr21 genes				
Hs.59471[a]	4.71	21q22.12[b]	7	23.51
Hs.596883[a]	4.84	21q22.3[b]	8	42.32
Hs.665969[a]	4.93	21q22.3[b]	14	35.08
Hs.661911[a]	4.93	21q22.3[b]	7	64.42
Hs.603678[a]	5.00	21q22.3[b]	8	57.32

Value: mean gene expression value normalized across all the pool samples; Data points: number of spots associated to an expression value for the locus; 'SD': standard deviation for the expression value expressed as a percentage of the mean. Over-expressed genes are sorted by decreasing gene expression value, under-expressed genes are sorted by increasing gene expression value. Analysis was performed using default parameters (see "Results" section). Full results are available as supplemental data (see text). [a] This gene is one of the five most over-/under-expressed, but the value is not statistically significant. This is either because this gene was present with other genes even in the 12,500 bp segment and its expression value does not prevail over the others, or because the segment containing the gene does not fulfil criteria for local over-/under-expression, or because its expression value is not associated to at least 5 data points. [b] Cytoband not available in Gene was derived from the UCSC Genome Browser (http://genome-euro.ucsc.edu/cgi-bin/hgGateway).

In the analysis of the heart vs. pool of tissues minus heart transcriptome map, regional differential expression of pool 'G' (32 heart samples) versus pool 'H' (629 samples) was investigated. Results obtained by analysis included 13 significantly over-/under-expressed segments: nine of them are over-expressed, four are under-expressed. The genome segment that has the highest statistically significant expression ratio is on chromosome 1 (1q43) (Table 13), including the over-expressed known genes *ACTN2* and *RYR2*, encoding respectively for actinin, alpha 2 and ryanodine receptor 2 (cardiac), and the EST cluster Hs.601529. The genome segment that has the lowest statistically significant expression ratio is on chromosome 4 (4p21.1) (Table 13), including the under-expressed known genes: *STATH*, encoding for statherin, *HTN3*, encoding for histatin 3, *HTN1*, encoding for histatin 1, *FDCSP*, encoding for follicular dendritic

cell secreted protein, and *SMR3B* encoding for submaxillary gland androgen regulated protein 3B.

Table 13 The genomic segments significantly over-/under-expressed in the whole heart (pool 'G') vs. pool of tissues minus heart (pool 'H') differential transcriptome map.

Whole heart vs. Pool of tissues minus heart differential transcriptome map					
Chr[a] and Location	**Segment Start[b]**	**Segment End[b]**	**Value 'G'/'H'**	**q-value**	**Genes in the segment[c]**
Over-expressed segments					
chr1 1q43	236,750,001	237,250,000	15.78	0.00076	***ACTN2*, Hs.601529, *RYR2***
chr1 1p36	11,750,001	12,250,000	14.41	0.0064	***NPPA*, *NPPB*, *MFN2***
chr14 14q12-q13	34,000,001	34,500,000	4.22	0.00014	***NPAS3*, *EGLN3*, Hs.664985**
chr4 4q35.1	186,500,001	187,000,000	3.08	0.00027	***SORBS2*, Hs.481342, Hs.658732**
Under-expressed segments					
chr4 4q21.1	70,750,001	71,250,000	0.30	0.000026	***STATH*, *HTN3*, *HTN1*, *FDCSP*, *SMR3B***
chr14 14q32.13	94,750,001	95,250,000	0.33	0.0029	***SERPINA1*, *SERPINA11*, *SERPINA5***
chr8 8p23.1	6,750,001	7,250,000	0.34	0.00005	***DEFA6*, *DEFA4*, *DEFA1*, *DEFA5***

Analysis was performed using default parameters (see "Results" section). Over-expressed segments are sorted by decreasing 'G'/'H' ratio, under-expressed segments are sorted by increasing 'G'/'H' ratio. In the 'Map' mode, TRAM displays UniGene EST clusters (with the prefix "Hs." in the case of *H. sapiens*) only if they have an expression value. For simplicity, some segments are not shown because they overlap with those highlighted in one of the listed regions. The complete results for these models are available as online supplemental data (see text). [a] Chromosome. The segment location cytoband was derived from that of the first mapped gene within the segment. [b] Segment Start/End: chromosomal coordinates for each segment. [c] Significantly over-/under-expressed genes as marked in the TRAM results.

At single gene level, an increase of more than 10 times was observed in all the first 129 loci, mostly including known genes with known cardiac function (see Table 14 in the text and Table IV in the Data Supplement. The last is available at: http://apollo11.isto.unibo.it/heart, where the Pool 'G' here is named 'A' and the Pool 'H' here is named 'B'; in the event of acceptance of the manuscript it will be available at: http://apollo11.isto.unibo.it/suppl/will be available at: http://apollo11.isto.unibo.it/suppl/).

We also observed that in this range of expression ratio, one chr21 locus is included: the EST cluster Hs.664516 (18.28-fold increase) (Table 14).

Among the genes with the lowest 'G'/'H' expression ratio, a 100-fold decrease was observed for the EST cluster Hs.663101 and for the known gene *ZG16B*, encoding for zymogen granule protein 16B (Table 14). Among the chr21 genes, the known gene *DSCR9*, encoding for Down syndrome critical region 9 (non-protein coding) has the lowest 'G'/'H' expression ratio, preceded by three EST clusters (Table 14).

Table 14 List of the five most over- or under-expressed genes (all significantly, with q<0.05, except genes with [a] note) in the whole heart (pool 'G') vs. pool of tissue minus heart (pool 'H') differential transcriptome map.

Whole heart vs. Pool of tissues minus heart differential transcriptome map								
Gene name	**Value 'G'**	**Value 'H'**	**Ratio 'G'/'H'**	**Location**	**Data points 'G'**	**Data points 'H'**	**SD as % of expression 'G'**	**SD as % of expression 'H'**
Over-expressed genes								
TNNI3[a]	4,648.05	23.33	199.26	19q13.4	26	568	76.26	207.39
Hs.603146[a b]	1,745.82	10.13	172.32	16q24.2	7	73	263.69	78.91
Hs.599556[a b]	1,666.76	10.62	156.91	9q22	7	73	263.38	87.26
Hs.374521[b]	1,362.16	10.52	129.43	Xq13.1	7	73	263.11	56.11
TNNT2	3,947.73	37.86	104.27	1q32	54	933	52.86	803.58
Over-expressed chr21 genes								
Hs.664516[a b]	773.63	42.33	18.28	21q22.11	7	73	258.34	65.20
ATP5O	2,205.50	382.56	5.77	21q22.11	34	1279	72.59	113.36
UMODL1	25.68	6.71	3.83	21q22.3	23	335	102.67	112.05
C21orf90[a]	82.59	22.84	3.62	21q22.3	25	720	411.67	69.30
ATP5J	2,254.25	648.88	3.47	21q21.2	36	575	84.88	77.51
Under-expressed genes								
Hs.663101[a b]	7.99	598.63	0.01	13q32.3	7	73	44.38	374.51
ZG16B	10.54	708.46	0.01	16p13.3	21	416	70.61	316.85
BPIFB1	6.89	456.21	0.02	20q11.21	25	462	87.51	401.41
Hs.605373[b]	5.12	316.85	0.02	15q15-q21.1	7	73	46.73	383.85
Hs.720020[b]	7.19	399.14	0.02	10q22.3	10	73	46.82	411.17
Under-expressed chr21 genes								
Hs.721038[a b]	6.40	81.31	0.08	21q22.3	6	355	68.88	114.43
Hs.663673[a b]	22.63	190.27	0.12	21q22.2	5	420	79.54	100.55
Hs.709790[a b]	34.58	265.04	0.13	21q22.11	5	425	156.19	105.62
DSCR9	5.39	38.12	0.14	21q22.13	24	405	77.30	80.44
TMPRSS2	13.97	96.85	0.14	21q22.3	58	1821	90.39	328.48

Value: mean gene expression value normalized across all the pool samples; Data points: number of spots associated to an expression value for the locus; SD: standard deviation for the expression value expressed as a percentage of the mean. Over-expressed genes are sorted by decreasing gene expression ratio, under-expressed genes are sorted by increasing gene expression ratio. Analysis was performed using default parameters (see "Results" section). Full results are available as supplemental data (see text). [a] This gene is one of the five most over-/under-expressed, but the value is not statistically significant. This is either because this gene was present with other genes even in the 12,500 bp segment and its expression value does not prevail over the others, or because the segment containing the gene does not fulfil criteria for local over-/under-expression, or because its expression value is not associated to at least 5 data points. [b] Cytoband not available in Gene was derived from the UCSC Genome Browser (http://genome-euro.ucsc.edu/cgi-bin/hgGateway).

4.1.11 Gene expression level in the heart and association to cardiac mutant phenotypes

Differential expression data were available in the TRAM 'G'/'H' table for 101 potassium channel encoding genes (named by a *KCN*- root). The certain or suspected association to a cardiac phenotype has been recorded (Table V in the Data Supplement, it is available at: http://apollo11.isto.unibo.it/heart, where the Pool 'G' here is named 'A' and the Pool 'H' here is named 'B'; in the event of the acceptance of the manuscript it will be available at: http://apollo11.isto.unibo.it/suppl/). It may be observed a clear clustering of gene associated to multiple cardiac phenotypes, when mutated, in the group of genes with the higher heart/non-

cardiac tissues expression ratio. In particular, following establishment of the arbitrary threshold of 1.5 ratio to define specific over-expression in the heart, 15 genes were expressed greater than 1.5 folds in the heart in comparison with non-cardiac tissues and 86 were expressed lower than 1.5 fold. Among the first group, 33.3% of the genes have been associated to known human arrhythmias (5 out of 15), but the association was seen only in 5.8% (5 out of 86) of the genes in the second group (p=0.006 by Fisher's exact test). A group of potassium receptor genes associated to non-cardiac, neurological phenotypes (*KCNB1* and *KCNA2*, epileptic encephalopathy, early infantile; *KCNMA1*, generalized epilepsy and paroxysmal dyskinesia; *KCNA1*, episodic ataxia/myokymia syndrome) has been found in the lower range of heart/non-cardiac tissues expression ratio (0.28-0.29). Searches in the previously reported quantitative transcriptome map for the human brain (Caracausi et al., 2014) revealed that these genes are instead actually over-expressed in the human brain in comparison to non-brain tissues, at a ratio of 3.2 (*KCNB1*), 2.8 (*KCNA2*), 2.2 (*KCNA1*) and 1.4 (*KCNMA1*).

4.1.12 Validation of TRAM map results through Real-Time RT-PCR

In order to confirm the results of our meta-analysis study, we performed "Real Time" RT-PCR experiments which are useful for validating the transcriptome maps of whole brain, cerebellum, cerebral cortex, hippocampus and whole heart. The primer pairs used to validate results are listed in Table 15.

Table 15 Primer pairs used to validate TRAM Maps.

Gene Symbol	**Gene Name**	**RefSeq RNA GenBank Accession No.**	**Primer pairs sequence (5′→ 3′)**[a]	**RT-PCR product size** (base pair)
BCYRN1	Brain cytoplasmic RNA 1	NR_001568	tagcgagaccccgttctccag gttgctttgagggaagttacgc	**82**
TUBA1B	Tubulin, alpha 1b	NM_006082	cctcgactcttagcttgtcgg aggcagtagagctcccagcag	**159**
SOD1	Superoxide dismutase 1, soluble	NM_000454	tagcgagttatggcgacgaag ggtacagcctgctgtattatctc	**186**
BEX5	Brain expressed, X-linked 5	NM_001012978	gtggtagaagctgacccctgag gctcctggtattcaccacctc	**226**
RCAN1	Regulator of calcineurin 1	NM_004414	ctggagcttcattgactgcgag gtgatgtccttgtcatacgtcc	**153**
GPCPD1	Glycerophos-phocholine phosphodiesterase GDE1 homolog (S. cerevisiae)	NM_019593	actcatggacctcagatctcg ggatcattggtatcatcaccc	**174**

OPRD1	Opioid receptor, delta 1	NM_000911	ctgggcaacgtgcttgtcatg catcaggtacttggcactctg	**141**
TBX18	T-box 18	NM_001080508	ctcgggggagacttggatgag ctgattctgataggcagtgacg	**247**
PABPN1L	Poly(A) binding protein, nuclear 1-like (cytoplasmic)	NM_001080487	aggagaaggtggaggctgacc gtccagagaacttgtcacacag	**141**
NPTX1	Neuronal pentraxin 1	NM_002522	gggcaaactttgcaatcgctc tctcctcggtgtcgttcctg	**193**
NUAK1	NUAK family, SNF1-like kinase, 1	NM_014840	tcaatgggagaccttaccgag tactctccgctgctgatttg	**139**
NTNG1	Netrin G1	NM_001113226	gaaagtgaaactcgatcctccg ctcgcatcacactcattattgc	**102**
ANKRD55	Ankyrin repeat domain 55	NM_024669	ggcttgaaggctgtgtgagtc gcagcatttgtgtgtgttgaggc	**228**
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	NM_002046	caacgaccactttgtcaagc ctgtgaggaggggagattca	**214**
NCDN	Neurochondrin	NM_001014839	ctgccacatcttcctcaacctc ccagggtggccacattagcag	**155**
SERPFIN1	Serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1	NM_002615	ggctgtctccaacttcggcta cggtgaatgatggattctgttcg	**150**
INADL	InaD-like (Drosophila)	NM_176877	ctcacacttcagcagtccatc ggaaccatctgtgaacactaac	**117**
MYL2	myosin, light chain 2, regulatory, cardiac, slow	NM_000432	gcgagtgaacgtgaaaaatgaag gaatggtttcctcagggtccg	**128**
NDUFA4	NDUFA4, mitochondrial complex associated	NM_002489	aacaaactgggtcccaatgatc ttgtgcggatgtggcttctgg	**151**
RYR2	ryanodine receptor 2 (cardiac)	NM_001035	gaaacagaacacacaggacagg gctggtcttcatactgtttccg	**109**
LRP5	low density lipoprotein receptor-related protein 5	NM_002335	cttcccgcacgagtatgtcag agctcatcatggactttccgc	**116**
SVOP	SV2 related protein homolog (rat)	NM_018711	tgcatttgcgcccgtgtatag acctcgaacactgtcccgatg	**174**

[a] Top: forward primer; bottom: reverse primer (for each gene).

Using criteria as described in the "Material and Method" section above, we selected 9 genes from the adult brain transcriptome map: *BCYRN1*, *TUBA1B*, *SOD1*, *BEX5*, *RCAN1*, *GPCPD1*, *OPRD1*, *TBX18* and *PABPN1L*. *BEX5* gene was chosen as reference gene. The *in vitro* observed gene expression ratios between each target gene and the reference gene are provided in Table 16.

Table 16 Selected genes to validate *in vitro* through Real-Time RT-PCR the whole adult brain, cerebellum, cerebral cortex, hippocampus and heart transcriptome maps.

Official Gene Symbol	Official Gene Name	EEV	SD	ER	Ct mean	OR
Brain						
BCYRN1	Brain cytoplasmic RNA 1	7,875.54	13.02	12.80	18.67	14.22
TUBA1B	Tubulin, alpha 1b	4,876.25	55.86	8.00	19.63	7.31
SOD1	Superoxide dismutase 1, soluble	1,727.39	55.37	2.80	20.12	5.21
BEX5	Brain expressed, X-linked 5	615.23	81.22	1.00	22.05	1.00
RCAN1	Regulator of calcineurin 1	212.66	218.52	0.34	24.10	0.33
GPCPD1	Glycerophosphocholine phosphodiesterase GDE1 homolog (S. cerevisiae)	63.29	92.61	0.10	26.08	0.084
OPRD1	Opioid receptor, delta 1	27.07	217.68	0.044	29.69	0.0068
TBX18	T-box18	9.65	61.70	0.015	29.92	0.0058
PABPN1L	Poly(A) binding protein, nuclear 1-like	4.57	8.77	0.007	34.09	0.00032
Cerebellum						
TUBA1B	Tubulin, alpha 1b	2,991.80	35.16	7.05	20.26	20.53
NPTX1	Neuronal pentraxin 1	1,081.87	86.51	2.55	22.53	4.26
NUAK1	NUAK family, SNF1-like kinase, 1	424.37	41.82	1.00	24.62	1.00
RCAN1	Regulator of calcineurin 1	142.60	166.56	0.34	23.41	2.31
GPCPD1	Glycerophosphocholine phosphodiesterase GDE1 homolog (S. cerevisiae)	167.03	95.59	0.39	23.32	2.46
NTNG1	Netrin G1	39.16	74.20	0.090	26.38	0.30
ANKRD55	Ankyrin repeat domain 55	17.19	52.41	0.041	31.34	0.010
TBX18	T-box 18	8.96	122.35	0.021	28.74	0.058
Cerebral cortex						
TUBA1B	Tubulin, alpha 1b	6,343.86	43.54	7.99	18.50	55.72
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	2,947.49	103.08	3.71	17.31	127.12
NUAK1	NUAK family, SNF1-like kinase, 1	793.97	62.54	1.00	24.30	1.00
NCDN	Neurochondrin	333.60	125.52	0.42	22.35	3.86
SERPINF1	Serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1	131.02	59.12	0.17	24.29	1.0070
RCAN1	Regulator of calcineurin 1	81.21	90.00	0.10	24.00	1.23
OPRD1	Opioid receptor, delta 1	42.71	80.72	0.13	27.77	0.090
INADL	InaD-like (Drosophila)	22.34	71.29	0.07	29.32	0.031
Hippocampus						
TUBA1B	Tubulin, alpha 1b	4,488.91	44.32	2.7	19.59	2.43
SOD1	Superoxide dismutase 1	1,647.81	28.92	1	20.87	1
NPTX1	Neuronal pentraxin 1	725.04	96.81	0.44	22.86	0.25
NUAK1	NUAK family SNF1-like kinase 1	522.06	110.21	0.32	24.44	0.084
SERPINF1	Serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1	255.58	283.85	0.16	24.36	0.089
NTNG1	Netrin G1	54.11	91.87	0.033	26.27	0.024
INADL	InaD-like (Drosophila)	24.57	57.20	0.015	28.49	0.005
Heart						
MYL2	myosin, light chain 2, regulatory, cardiac, slow	10,216.56	41.46	28.34	16.25	111.43

NDUFA4	NDUFA4, mitochondrial complex associated	3,706.51	53.98	10.28	18.47	23.92
RYR2	ryanodine receptor 2 (cardiac)	1,213.96	101.7	3.37	23.16	0.93
SERPINF1	Serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1	360.44	66.62	1	23.05	1
RCAN1	regulator of calcineurin 1	154.34	142.4	0.43	26.97	0.07
LRP5	low density lipoprotein receptor-related protein 5	55.83	198.22	0.15	26.29	0.11
SVOP	SV2 related protein homolog (rat)	19.86	58.48	0.06	32.49	0.0014

In bold the gene chosen as reference, underlined genes having a high standard deviation for each map. From left to right: official gene symbol of selected gene; official full gene name; expected expression value (EEV), i.e. expression value as provided by TRAM software; standard deviation (SD) as percentage of expression; expected ratio (ER) between reference and target gene expression value; threshold cycle (Ct) provided by BioRad CFX Manager Software 2.1 manually positioning the threshold line; observed ratio (OR) determined between each target gene and the reference gene using the delta Ct (ΔCt) method, according to the formula: $2^{\Delta Ct} = 2^{\text{Ctref - Cttarget}}$.

The correlation between the observed and expected gene expression ratios as calculated by bivariate analysis was statistically highly significant (Pearson correlation coefficient= 0.98 and p-value= 0) (Figure 3a).

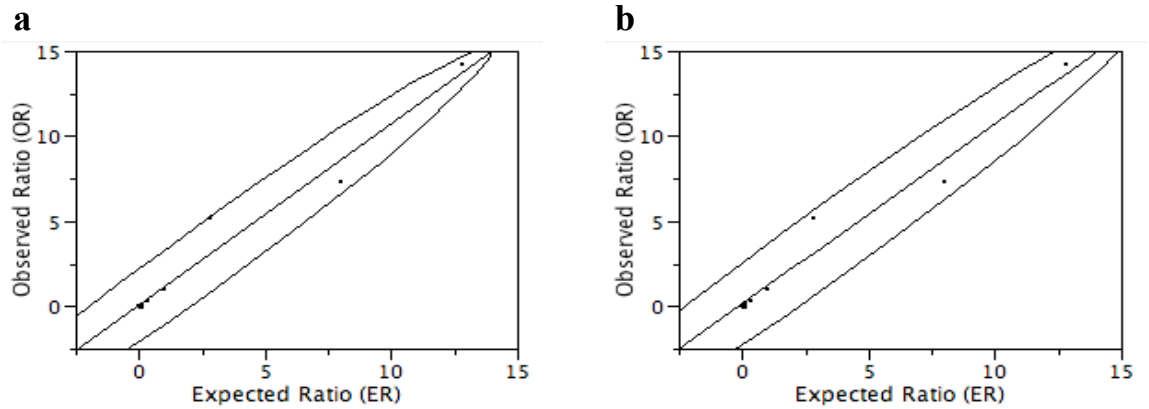


Figure 3 Bivariate analysis between observed (Real Time RT-PCR) and expected (TRAM) expression ratio in brain of selected genes (Table 16). **a** The fit line and a 95% bivariate normal density ellipse are shown; Pearson correlation coefficient is 0.98 and p-value<0.0001. **b** The fit line and a 95% bivariate normal density ellipse determined after the exclusion from the statistical analysis of those genes having a high standard deviation (SD as percentage of expression >95) of expected expression value (Table 16). This step did not affect the correlation between the two variables, indeed Pearson correlation coefficient is 0.98 and p-value 0.0001

From the whole cerebellum transcriptome map we selected the 7 genes: *TUBA1B*, *NPTX1*, *NUAK1*, *RCAN1*, *GPCPD1*, *NTNG1*, *ANKRD55* and *TBX18*. *NUAK1* gene was chosen as reference gene. We determined the *in vitro* observed expression ratio between each target gene and the reference gene (Table 16) as discussed above and we determined the correlation through

bivariate analysis. The result was again statistically highly significant (Pearson correlation coefficient=0.97 and p-value<0.0001) (Figure 4a).

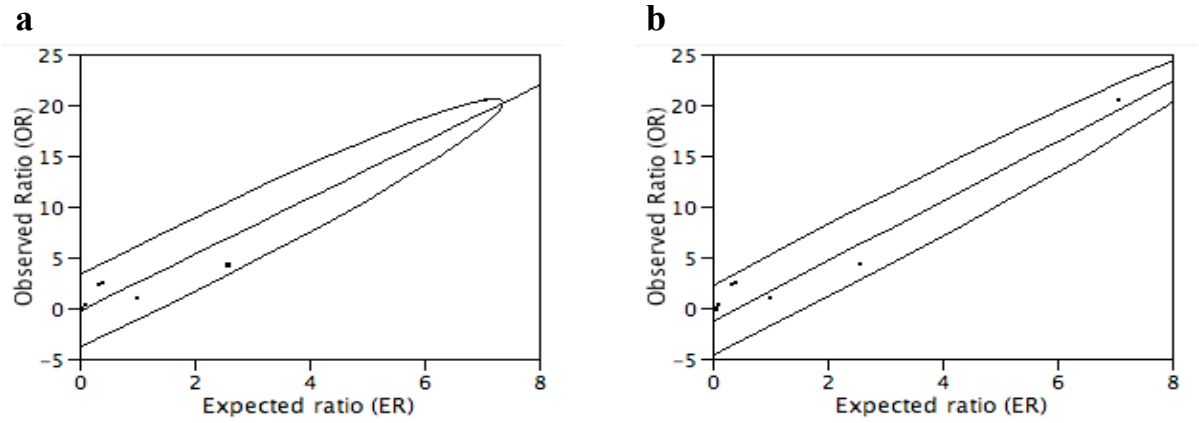


Figure 4 Bivariate analysis between observed (Real Time RT-PCR) and expected (TRAM) expression ratio in cerebellum of selected genes (Table 16). **a** The fit line and a 95% bivariate normal density ellipse are shown; Pearson correlation coefficient is 0.97 and p-value<0.0001. **b** The fit line and a 95% bivariate normal density ellipse determined after the exclusion from the statistical analysis of those genes having a high standard deviation (SD as percentage of expression >95) of expected expression value (Table 16). This step improved the correlation between the two variables, indeed Pearson correlation coefficient is 0.98 and p-value 0.0019.

From the whole cerebral cortex transcriptome map we selected the 8 genes: *TUBA1B*, *GAPDH*, *NUAK1*, *NCDN*, *SERPINF1*, *RCAN1*, *OPRD1* and *INADL*. *NUAK1* gene was chosen as reference gene. We determined the *in vitro* observed expression ratio between each target gene and the reference gene (Table 16) as discussed above and we performed the bivariate analysis between the expected and the observed ratios. The correlation between the two variables was not statistically significant (Figure 5a). When we attempted to exclude those genes that have an expected expression value with a high standard deviation (SD as percentage of Expression > 95) from the statistical analysis, assuming that a high standard deviation implies a greater variability in the observed data, the correlation between the two variables became statistically highly significant (Pearson correlation coefficient=0.99 and p-value<0.0001) (Figure 5b).

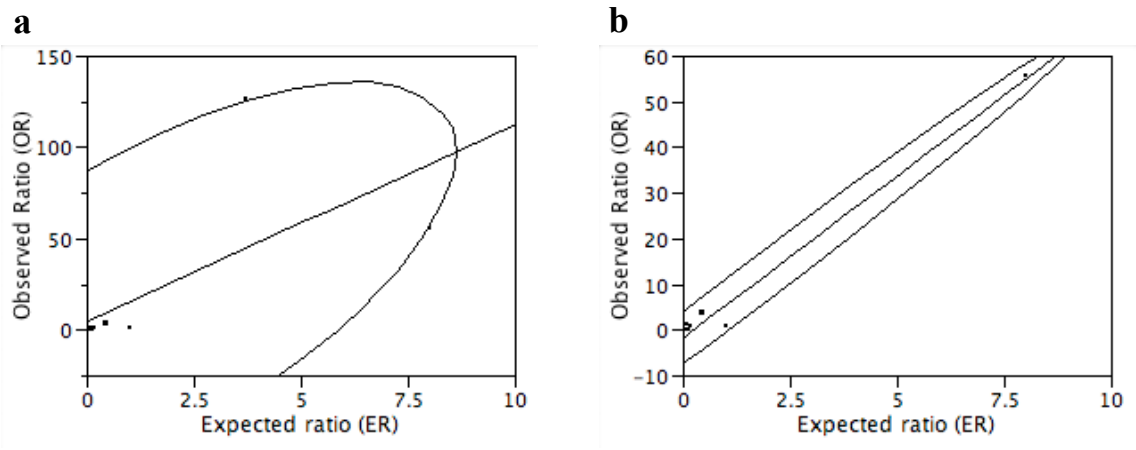


Figure 5 Bivariate analysis between observed (Real Time RT-PCR) and expected (TRAM) expression ratio in cerebral cortex of selected genes (Table 16). **a** The fit line and a 95% bivariate normal density ellipse are shown. The correlation between the two variables was not statistically significant, the Pearson correlation coefficient is 0.65 and p-value 0.0762. **b** The fit line and a 95% bivariate normal density ellipse determined after the exclusion from the statistical analysis of those genes having a high standard deviation (SD as percentage of expression >95) of expected expression value (Table 16). This step showed an increase of the correlation between the two variables, indeed the Pearson correlation coefficient is 0.99 and p-value<0.0001.

Applying this criterion to the previous statistical analysis, we observed that for the adult brain data there was no change (Figure 3b), for the whole cerebellum data occurred an increase of the Pearson correlation coefficient to 0.98, while the p-value became 0.0019 (Figure 4b). Despite these variations, all correlations remained significant.

We selected 7 genes from the hippocampus transcriptome map: *TUBA1B*, *SOD1*, *NPTX1*, *NUAK1*, *SERPINF1*, *NTNG1*, *INADL*. *SOD1* gene was chosen as reference gene. The primer pairs used to validate results are listed in Table 15. The *in vitro* observed gene expression ratios between each target gene and the reference gene are provided in Table 16. The correlation between the observed and expected gene expression ratios as calculated by bivariate analysis was statistically highly significant: Pearson correlation coefficient is 0.99 and p-value<0.0001 (Figure 6).

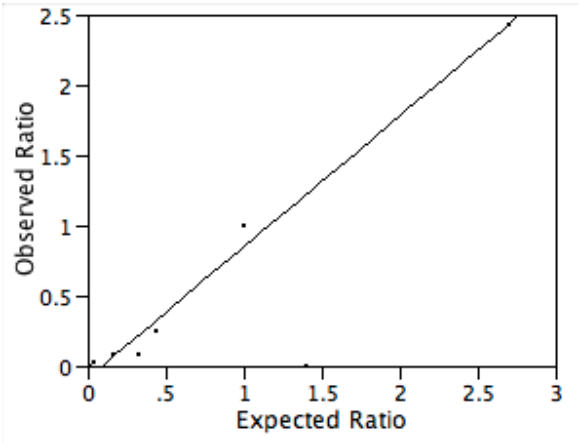


Figure 6 Bivariate analysis between observed (Real-Time RT-PCR) and expected (TRAM) expression ratios of selected genes in the hippocampus transcriptome map (Table 16). The fit line is shown; Pearson correlation coefficient is 0.99 and p-value<0.0001.

We selected 7 genes from the heart transcriptome map: *MYL2*, *NDUFA4*, *RYR2*, *SERPINF1*, *RCAN1*, *LRP5*, *SVOP*. *SERPINF1* gene was chosen as reference gene. The primer pairs used to validate results are listed in Table 15. The *in vitro* observed gene expression ratios between each target gene and the reference gene are provided in Table 16. The correlation between the observed and expected gene expression ratios as calculated by bivariate analysis was statistically highly significant: Pearson correlation coefficient is 0.98 and p-value<0.0001 (Figure 7).

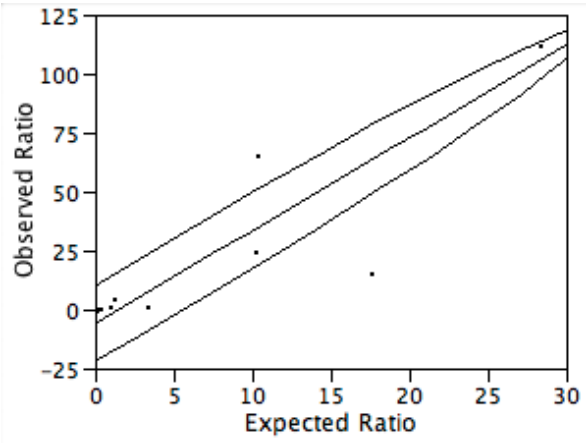


Figure 7 Bivariate analysis between observed (Real-Time RT-PCR) and expected (TRAM) expression ratios of selected genes in the whole heart transcriptome map (Table 16). The fit line is shown; Pearson correlation coefficient is 0.98 and p-value<0.0001.

4.1.13 Housekeeping gene search

In the pool of normal tissues minus brain transcriptome map, using the lower SD as a percentage of the mean value (≤30), two EST clusters (Hs.714416 and Hs.728191) and only one known gene (*POM121C*, encoding for the transmembrane nucleoporin C) were retrieved; while,

using the higher SD as a percentage of the mean value at ≤40, 29 genes common to all pool tissues fulfilled the selected criteria. By applying functional enrichment analysis with the "ToppGene Suite" Gene Ontology tool (Chen et al., 2009), we found that the statistically significant (p-value≤0.05) enriched function is structural constituent of ribosome (GO:0003735), associated to the *RPS17L*, *MRPL18* and *RPS18* genes. Ten out of 29 genes resulted "not found": they are the EST clusters.

In the adult brain, using the lower SD as a percentage of the mean value (≤30), one EST cluster (Hs.705664) and only one known gene (*FUNDC1*, encoding for an integral mitochondrial outer-membrane protein) were retrieved; while, using the higher SD as a percentage of the mean value at ≤40, 15 genes fulfilled the selected criteria. By applying functional enrichment analysis, we found that the statistically significant (p-value≤0.05) enriched function is unfolded protein binding (GO:0051082), associated to the *PDRG1* and *TTC1* genes. Six out of 15 genes resulted "not found": they are the EST clusters.

The same thing was done for the foetal brain transcriptome map. In this case, using the lower SD as a percentage of the mean value (≤30), 40 genes including some ribosomal proteins (*MRPS24*, *MRPS35* and *RPS18*) were retrieved; while, using the higher SD as percentage of the mean value at ≤40, 229 genes came out by the selection. The statistically significant (p-value≤0.05) enriched function is a structural constituent of ribosome (GO:0003735), associated to the *MRPL18*, *RPL23*, *RPL4*, *RPS17L*, *MRPL51*, *RPS13*, *MRPL14*, *MRPS24*, *RPL24*, *RPS24*, *RPS18* and *MRPL32* genes.

Analyzing common results among the three maps: *RPS17L*, encoding for S17-like ribosomal protein, is the only gene common to all three maps; the EST cluster Hs.714416 is common to the pool of tissues and adult brain transcriptome maps; *PDRG1*, encoding for p53 and DNA-damage regulated 1, *TCEAL8*, encoding for transcription elongation factor A (SII)-like 8, and *TIMM9*, encoding for translocase of inner mitochondrial membrane 9 homolog (yeast), are common to the foetal brain and brain transcriptome maps; *ACTG1*, encoding for actin, gamma 1, *EMC4*, encoding for ER membrane protein complex subunit 4, *MRPL18*, encoding for mitochondrial ribosomal protein L18, and *YTHDF1*, encoding for YTH domain family, member 1, are common to the pool of tissues and foetal brain transcriptome maps.

These results are provided as Supplementary Table 15, available at: http://apollo11.isto.unibo.it/suppl/2014_Caracausi/.

In the hippocampus transcriptome map, the search for housekeeping genes retrieved 241 loci, including many known genes and uncharacterized loci (mostly EST clusters), fulfilling the selected criteria (Supporting Information Table S5, available at: http://apollo11.isto.unibo.it/suppl/2015_Caracausi/).

Within this list, *TBCB* (tubulin folding cofactor B beta) was found to be the gene with the highest number of both data points (n=93) and samples (all 41 samples covered), suggesting the

reliability of its expression values (mean=570.15, SD expressed as a percentage of the mean=17.76).

By applying functional enrichment analysis we found that there are fifteen statistically significant (p-value≤0.05) enriched functions (Supporting Information Table S6, available at: http://apollo11.isto.unibo.it/suppl/2015_Caracausi/). The first three are: RNA binding (GO:0003723), poly(A) RNA binding (GO:0044822) and oxidoreductase activity (GO:0016491) (Supporting Information Table S6, available at: http://apollo11.isto.unibo.it/suppl/2015_Caracausi/). Seventy out of 241 genes resulted "not found", they are: the EST clusters; an uncharecterized locus *LOC100506365*; two known genes, *GTDC2* and *NIM1*, encoding respectively for protein O-linked mannose N-acetylglucosaminyltransferase 2 (beta 1,4-) and for NIM1 serine/threonine protein kinase.

In the heart transcriptome map, the search for housekeeping genes with an SD as percentage of the mean value ≤30 retrieved 2 loci, including the known gene *MRPL51*, encoding for mitochondrial ribosomal protein L51, and the EST cluster Hs.465405, fulfilling the selected criteria (Table 17). The search for housekeeping genes with an SD as percentage of the mean value ≤40 retrieved 7 loci (Table 17).

Table 17 Predicted housekeeping genes from the whole heart transcriptome maps using two different SD as percentage of the mean value (≤30 and ≤40).

Gene name	Value 'G'	Chr	Location	Data points 'G'	SD as % of expression 'G'
Whole heart housekeeping genes having SD as % of the mean value ≤30					
Hs.465405[a]	121.91	chr1	1p36.13	19	22.92
MRPL51	715.22	chr12	12p13.3-p13.1	18	27.31
Whole heart housekeeping genes having SD as % of the mean value ≤40					
Hs.465405[a]	121.91	chr1	1p36.13	19	22.92
MRPL51	715.22	chr12	12p13.3-p13.1	18	27.31
RRAGC	126.38	chr1	1p34	32	33.07
TRMT10C	103.35	chr3	3q12.3	119	37.41
TRMT112	422.61	chr11	11q13.1	31	39.05
ECHS1	764.25	chr10	10q26.2-q26.3	30	39.56
PIK3C2B	106.77	chr1	1q32	35	39.70

Genes are sorted in ascending order of SD as percentage of the mean value. [a] Cytoband not available in Gene was derived from the UCSC Genome Browser (http://genome-euro.ucsc.edu/cgi-bin/hgGateway).

4.2 DS AMKL transcriptome

4.2.1 AMKL gene expression literature search

A general search using the acronym "AMKL" retrieved 157 articles, 6 of them describe gene expression profiling experiments (Klusmann et al., 2010a; Klusmann et al., 2010b; Lightfoot et al., 2004; McElwaine et al., 2004; Bourquin et al., 2006; Ge et al., 2006).

No additional pertinent item was retrieved using the expression described in the Methods section and including the MeSH Term "Leukemia, Megakaryoblastic, Acute".

4.2.2 AMKL gene expression database search

The Gene Expression Omnibus (GEO) (Barrett and Edgar, 2006) search allowed the retrievement of three additional works describing data possibly useful for meta-analysis (Yagi et al., 2003; Radtke et al., 2009; Kikushige et al., 2010). The lack of inclusion of these works in the literature search was due to failure of using the "AMKL" acronym and assigning the MeSH Term "Leukemia, Megakaryoblastic, Acute" during the PubMed indexing process (the more general term "Leukemia, Myeloid, Acute" was used).

The more general search using the expression "Down Syndrome"[MeSH] AND "Homo sapiens"[Organism] allowed the addition of one further work (Tomasson et al., 2008). This work analyzed several types of AML samples and did not explicitly mention AMKL or AML M7 in both PubMed and GEO databases.

No further pertinent works related to AMKL were identified by ArrayExpress database (Sarkans et al., 2005) search.

Several datasets for normal MK cells global gene expression profile fulfilling the selection criteria were obtained from the works (Nover-shtern et al., 2011) (GEO, 7 samples) and (Felli et al., 2010) (ArrayExpress, 1 sample), in addition to the 4 sample series identified in the first report of the TRAM software (Lenzi et al., 2011) and obtained from different works (Tenedini et al., 2004; Fuhrken et al., 2007; Giammona et al., 2006; Ferrari et al., 2007), for a total of 19 datasets related to human normal MK cells.

4.2.3 Dataset building

Of the 10 works related to DS or non-DS AMKL retrieved as above described, 7 were considered for the meta-analysis (Table 18). It was not possible to obtain raw data from the Authors of (Lightfoot et al., 2004), while the only sample of AML M7 described in (Kikushige et al., 2010) was related to "Leukemic Stem Cells" cell type and the two AML M7 reported by Tomasson et al. (2008) were obtained from elderly patients. Data from (Ge te al., 2006) were kindly provided by Drs. Jeffrey Taub and Yubin Ge.

At the end, DS AMKL sample pool 'A' included 43 datasets, while non-DS AMKL sample pool 'B' was composed of 45 datasets. A TMD dataset pool 'C' was constructed starting from 20 samples described in some of the DS AMKL related articles (Klusmann et al., 2010b;

McElwaine et al., 2004; Bourquin et al., 2006). Age and sex data were available for 29 out of 43 DS AMKL patients (mean age: 20 months; 11 males and 18 females), for 26 out of 45 non-DS AMKL patients (mean age: 19 months; 19/7 males/females) and for 9 out of 20 TMD patients (mean age: 8 days; 7/2 males/females). *GATA1* mutations giving rise to GATA1s were present in all DS AMKL and TMD samples, and not in non-DS AMKL samples, considering all samples for which this information was provided. Sample identifiers and main sample features are listed in Table 18 and Additional file 1 (available at: http://apollo11.isto.unibo.it/suppl).

Two pools were constructed from the normal MK related dataset selected: pool 'D' included all available MK samples, while pool 'E' was a subset including only CB-derived MK cells (Table 18 and Additional file 1, the last is available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/).

Table 18. Main features of the samples used in TRAM analyses.

Study ID	Sample ID	Sample type	Platform	Microarray	Spots	References
Pool 'A' **DS AMKL** (25,955 mapped loci following analysis by TRAM) **(n=43)**						
A1...A3 (n=3)	GSM491372...4	BM Sorted leukemic blasts	GPL570	Affymetrix U133 Plus 2.0	54,675	Klusmann 2010b
A4...A25 (n=22)	GSM94245, GSM94272...92	BM or PB	GPL96	Affymetrix U133A	22,283	Bourquin 2006
A26...A31 (n=6)	GSM417985...90	BM or PB Sorted leukemic blasts	GPL570	Affymetrix U133 Plus 2.0	54,675	Klusmann 2010a
A32...A38 (n=7)	E-MEXP-72*	BMMC or PBMC	A-AFFY-33.adf.txt*	Affymetrix U133A	22,283	McElwaine 2004
A39...A43 (n=5)	/	BMMC or PBMC	GPL96	Affymetrix U133A	22,283	Ge 2006
Pool 'B' **non-DS AMKL** (26,045 mapped loci following analysis by TRAM) **(n=45)**						
B1-B2 (n=2)	GSM491370-1	BM Sorted leukemic blasts	GPL570	Affymetrix U133 Plus 2.0	54,675	Klusmann 2010b
B3...B23 (n=21)	GSM94221-4-5, GSM94227...32, GSM94234-5-7-8, GSM94240-2-3-8, GSM94256-9, GSM94261-2	BM or PB	GPL96	Affymetrix U133A	22,283	Bourquin 2006
B24...B28 (n=5)	GSM39832-5, GSM39842-4, GSM39863	PBMC or BMMC	GPL8300	Affymetrix U95 Version 2	12,625	Yagi 2003
B29...B35 (n=7)	GSM361502...4, GSM361506...8, GSM361510	BM or PB	GPL96	Affymetrix U133A	22,283	Radtke 2009
B36...B40 (n=5)	GSM417991...5	BM or PB Sorted leukemic blasts	GPL570	Affymetrix U133 Plus 2.0	54,675	Klusmann 2010b

B41...B45 (n=5)	/	BMMC or PBMC	GPL96	Affymetrix U133A	22,283	Ge 2006
Pool 'C' **TMD** (25,955 mapped loci following analysis by TRAM) **(n=20)**						
C1...C3 (n=3)	GSM491375...7	BM Sorted leukemic blasts	GPL570	Affymetrix U133 Plus 2.0	54,675	Klusmann 2010b
C4...C11 (n=8)	GSM94293...9 GSM94300	BM or PB	GPL96	Affymetrix U133A	22,283	Bourquin 2006
C12...C20 (n=9)	E-MEXP-72*	BMMC or PBMC	A-AFFY-33.adf.txt*	Affymetrix U133A	22,283	McElwaine 2004
Pool 'D' **MK** (26,372 mapped loci following analysis by TRAM) **(n=19)**						
D1-D2 (n=2)	GSM321577-8	MK (BM) (subj=pool)	GPL96	Affymetrix U133A	22,283	Ferrari 2007
D3...D6 (n=4)	GSM112277-8, GSM112291-2	MK (PB) (subj=1, rep. 1)	GPL887	Agilent 1A	22,575	Giammona 2006
D7-D8 (n=2)	GSM15648, GSM8649	MK (BM) (subj=6)	GPL96	Affymetrix U133A	22,283	Tenedini 2004
D9...D11 (n=3)	GSM88014-22-34	MK (PB) (subj=1)	GPL887	Agilent 1A	22,575	Fuhrken 2007
D12...D18 (n=7)	GSM609746...52	CB	GPL4685	Affymetrix HT-HG_U133A	22,944	Nover-shtern 2011
D19 (n=1)	E-MEXP-2146*	CB	A-AFFY-44.adf.txt	Affymetrix U133 Plus 2.0	54,675	Felli 2010
Pool 'E' = D12...D19 **CB MK** (25,577 mapped loci following analysis by TRAM) **(n=8)**						

Samples selected for the meta-analysis of gene expression profiles in DS AMKL (pool 'A'), non-DS AMKL (pool 'B'), TMD (pool 'C'), and megakaryocytic cells (pool 'D' and 'E'). All Sample IDs and Platforms IDs are related to GEO database, other than codes marked with * (ArrayExpress database). Sample type: BM, bone marrow; PB, peripheral blood; BMMC, bone marrow mononuclear cells; PBMC, peripheral blood mononuclear cells; MK, megakaryocytic/megakaryoblast cells, obtained by *in vitro* differentiation of CD34+ cells; CD34+, undifferentiated CD34+ cells; BM, CB or PB: CD34+ cells derived from bone marrow, cord blood or peripheral blood, respectively. subj=number of subjects from which the sample was derived (in some cases, where subj=pool, the exact number of subjects included in a pool was not available). rep.=biological replicate. **Microarray:** U133A: Affymetrix Human Genome U133A Array; 1A: Agilent-012097 Human 1A Microarray (V2) G4110B; 22k A: Agilent Human oligo 22k A; HG-Focus: Affymetrix Human HG-Focus Target Array.
Details about Sample identifiers and main sample features are listed in Additional file 1 (available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/).

4.2.4 Transcriptome differential maps

Datasets were loaded into TRAM and analyzed obtaining 8 tanscriptome maps: DS AMKL (pool 'A') vs. non-DS AMKL (pool 'B'); DS AMKL (pool 'A') vs. normal MK (pool 'D'); non-DS AMKL (pool 'B') vs. normal MK (pool 'D'); DS AMKL (pool 'A') vs. normal CB MK (pool 'E') cells; non-DS AMKL (pool 'B') vs. normal CB MK (pool 'E'); DS AMKL (pool 'A') vs. TMD (pool 'C'); TMD (pool 'C') vs. normal MK (pool 'D'); TMD (pool 'C') vs. normal CB MK (pool 'E').

For each comparison between two cell types by TRAM, we describe below or in the corresponding Figures or Tables the total of data points analyzed for each cell type, i.e. gene expression values for all human mapped loci following intra- and inter-sample normalization (Lenzi et al., 2011); the number of loci for which the comparison between the two conditions was possible due to the presence of values for those loci in both sample pools considered; the number and the gene content of each genomic segment containing at least three over- or under-expressed genes and found to be statistically significantly over- or under-expressed in the comparison between the two tissues. Each genomic segment was identified among the 12,373 segments generated using the default window of 500,000 bp with a sliding window of 250,000 bp and following removal of overlapping segments with similar gene content.

The description of the gene name corresponding to all gene symbols cited here in the text, Figures or Tables is given in the Additional file 2, available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/. We performed a PubMed search for the most relevant over- or under-expressed genes using gene symbol or gene description along with MeSH terms related to MK or MK progenitor cells, thrombopoiesis, AMKL, platelets.

Detailed results for each map are provided below, and are also available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/.

The absolute (not differential) expression values and maps for each cell type (not compared to another cell type) are also available in the complete sets of results at available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/, but are not discussed here because they include typical housekeeping genes whose over-expression is no longer evident when compensated by the corresponding housekeeping genes in the compared cell type.

4.2.5 Transcriptome map comparison of DS AMKL vs. non-DS AMKL

We first analyzed regional differential expression of pool 'A' (43 DS AMKL samples) versus pool 'B' (45 non-DS AMKL samples) (Table 18). A total of 1,061,761 data points from the pool 'A' and 1,084,700 data points from the pool 'B' were included in the analysis. An 'A'/'B' ratio value was determinable for 25,954 loci having values both in 'A' and 'B' pools (Additional file 3, available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/). The main results are shown in Figure 8.

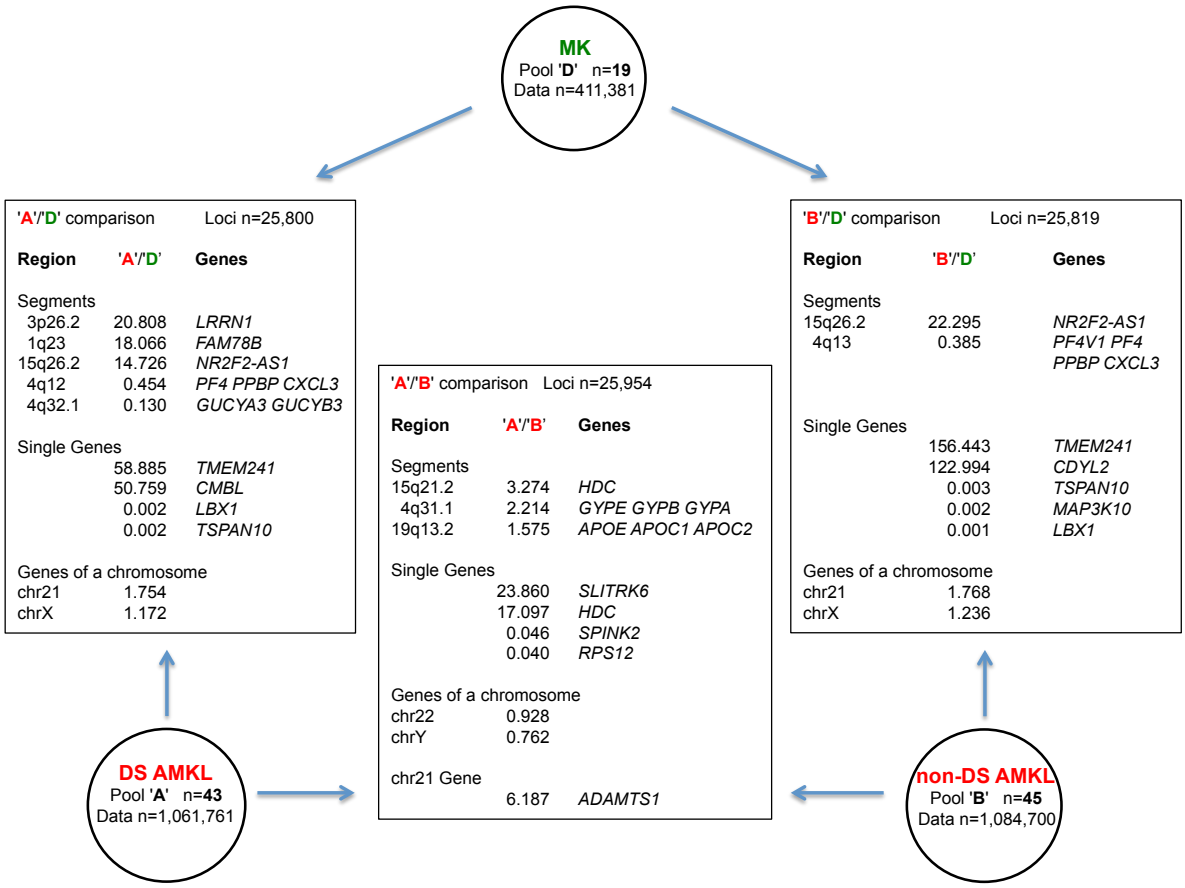


Figure 8 Main results of DS-AMKL vs. MK, DS AMKL vs. non-DS AMKL, non-DS AMKL vs. MK comparisons. For each comparison the number of loci analyzed, the most over- or under-expressed segments and single genes, and the highest and lowest median expression ratios for all the genes located in the same chromosome are indicated.

Results obtained by the analysis included 3 significantly non-overlapped over-expressed segments (Table 19). The highest expression ratio between DS AMKL and non-DS AMKL (3.27) was observed in a segment on chromosome 15 (15q21.2), including the known gene *HDC* (encoding for histidine decarboxylase, which converts L-histidine to histamine). The second segment with the highest expression was located on chromosome 4 (4q31.1) and contained over-expressed genes such as *GYPE*, *GYPB* and *GYPA*, encoding for glycophorin E, B and A (MNS blood group) respectively. The third over-expressed segment spans the cluster of apolipoprotein encoding genes on chromosome 19 (19q13.2).

Table 19 Genomic segments significantly over- or under-expressed in the DS AMKL vs. non-DS AMKL differential transcriptome map.

Chr and Location[a]	Segment Start[b]	Segment End[b]	'A'/'B' Ratio	q-value	Genes in the segment[c]
Over-expressed segments					
chr15 15q21.2	50,250,001	50,750,000	3.27	0.00233	***HDC*, Hs.656448, Hs.660869**[d]
chr4 4q31.1	144,750,001	145,250,000	2.21	0.00024	***GYPE*, *GYPB*, *GYPA***
chr19 19q13.2	45,000,001	45,500,000	1.58	0.00461	***APOE*, *APOC1*, *APOC2***

Data refer to the following comparisons: DS AMKL (pool 'A') vs. non-DS AMKL (pool 'B'). Analysis was performed using default parameters (see Methods section). Segments are sorted by decreasing 'A'/'B' ratio. In the "Map" mode, TRAM displays UniGene EST clusters (with the prefix "Hs." in the case of *Homo sapiens*) only if they have an expression value. Some segments are not shown for simplicity because they are over-lapping with those highlighted in one of the listed regions. The complete results for these models are available as on line additional material. [a] Chr: chromosome. The segment location cytoband was derived from that of the first mapped gene within the segment. [b] Segment Start/End: chromosomal coordinates for each segment. [c] Significantly over-/under-expressed genes as marked in the TRAM results. [d]This UniGene cluster contains EST with at least one Alu sequence, according to Repeat Masker (http://www.repeatmasker.org/). For this reason, it can not be excluded that its over-expression is related to unspecific hybridization by Alu-containg probes.

At single gene level, a fold increase higher than 5 was observed in all of the first 20 genes with the greatest expression ratios of DS AMKL vs. non-DS AMKL samples (Table 20 and Additional file 3, the last is available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/). In particular, a 24-fold increase was observed for *SLITRK6* gene, encoding for a membrane protein to date described as similar to receptor for BDNF (brain-derived neurotrophic factor) and predominantly expressed in neural tissues. Among the genes with the lower 'A'/'B' expression ratios a 169.5-fold decrease was observed for a UniGene EST cluster, Hs.355689.

Table 20 List of the five most over- or under-expressed genes (all significantly, with q<0.05) in the DS AMKL vs. non-DS AMKL transcriptome map.

Gene	Value 'A'	Value 'B'	'A'/'B' Ratio	Location	Data points		SD	SD
					'A'	'B'	'A'	'B'
Over-expressed genes								
SLITRK6	465.5	19.5	23.860	13q31.1	27	21	113.0	79.6
HDC	1,119.4	65.5	17.097	15q21-q22	36	45	98.8	249.4
ZNF587B	615.0	39.3	15.643	19q13.43[b]	43	40	427.6	32.2
SOSTDC1	111.7	8.6	12.957	7p21.1	43	45	176.5	108.4
LOC100287628[a]	1,381.3	107.7	12.820	16p13.2	9	7	80.2	56.8
Under-expressed genes								
Hs.587427[a]	19.8	407.8	0.049	7p15.2	18	14	28.9	353.3
SPINK2	13.1	285.1	0.046	4q12	36	45	81.8	115.0
Hs.602709[a]	15.8	370.2	0.043	11q13.2[b]	9	7	33.1	256.9
RPS12	20.7	519.1	0.040	6q23.2	43	50	113.7	213.0
Hs.355689[a]	3.7	624.4	0.006	18q11.2[b]	9	7	47.3	168.4

Data refer to the following comparisons: DS AMKL (pool 'A') vs. non-DS AMKL (pool 'B'). Value: mean gene expression value normalized across all the pool samples; Data points: number of spots associated to an expression value for the locus; SD: standard deviation for the expression value expressed as percentage of the mean. Full results available as additional material (see text). [a] The segment window contains more than one gene, but the significance is assumed to be maintained because the expression value of this over- or under-expressed gene prevails over the others. [b] Cytoband not available in Gene was derived from the UCSC Genome Browser (University of California Santa Cruz (UCSC) Genome Browser. (http://genome-euro.ucsc.edu/cgi-bin/hgGateway).

At chromosomal level, we calculated (in the TRAM "Chr" table) the median 'A'/'B' expression ratio for all the genes located in the same chromosome. The highest ratios were near to 1 (0.93 for chr22, 0.92 for chrX and chr21, 0.91 for chr19); other values were in the range from 0.90 (chr17 and chr12) to 0.76 (chrY).

We performed two additional transcriptome maps to investigate specifically sex-biased gene expression patterns (data not shown, results may be regenerated by the user by excluding/including or reimporting samples on the basis of data provided in Additional file 1, available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/): in particular, we compared male (pool 'A.1', n=11) vs. female DS AMKL cells (pool 'A.2', n=18). These datasets are derived from the samples for which the knowledge about the sex of the sample donor was available. The results showed a significant statistical correlation of data between male and female gene expression data (r=0.99, p-value<0.0001), showing a large overlap of results between the two transcriptome maps, with the exception of single genes with a well known sex-biased expression pattern. For example, *XIST*, which is specifically activated in female cells to start the X-inactivation process, turns out to be the most differentially expressed gene between female (value=402.60) and male (value=12.20) DS AMKL cells (ratio=33).

4.2.6 Transcriptome map comparison of DS AMKL or non-DS AMKL vs. normal MK

Regional differential expression of pool 'A' (43 DS AMKL samples) or pool 'B' (45 non-DS AMKL samples) versus pool 'D' (19 normal MK cell samples, 411,381 data points) (Table 18) was investigated. An 'A'/'D' ratio value was determinable for 25,800 loci (Additional file 4, available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/). The main results are shown in Figure 8.

For what DS AMKL samples are concerned, results included 5 significantly differentially expressed segments in DS AMKL cells, 3 over- and 2 under-expressed (Table 21). The highest expression ratio (20.81) between DS AMKL cells and normal MK was observed in the segment at coordinates 3,500,001-4,000,000 on chromosome 3, including the known gene *LRRN1*, encoding for a type I transmembrane protein. The second segment with highest expression ratio (18.07) was located on chromosome 1 (1q23) and contained *FAM78B* (family with sequence similarity 78, member B). The third segment was on chromosome 15 and included *NR2F2-AS1*, a non-coding RNA. The first significantly under-expressed segment (4q32.1) includes genes encoding for subunits of soluble guanylate cyclase (*GUCY1A3* and *GUCY1B3*), while the second spans the cluster of MK specific genes located on chromosome 4 (4q12-q21).

Table 21 Genomic segments significantly over- or under-expressed in the DS AMKL vs. MK differential transcriptome map.

Chr and Location[a]	Segment Start[b]	Segment End[b]	'A'/'D' Ratio	q-value	Genes in the segment[c]
Over-expressed segments					
chr3 p26.2	3,500,001	4,000,000	20.81	0.00006	**Hs.241414, *LRRN1*, Hs.128128**
chr1 1q23	166,000,001	166,500,000	18.07	0.00004	**Hs.22930, *FAM78B*, Hs.662048**
chr15 15q26.2	96,750,001	97,250,000	14.73	0.00001	**Hs.677040, Hs.661950, *NR2F2-AS1*, Hs.592015**
Under-expressed segments					
chr4 4q12-q21	74,750,001	75,250,000	0.45	0.00079	***PF4*, *PPBP*, *CXCL3***
chr4 4q32.1	156,250,001	156,750,000	0.13	0.00012	***GUCY1A3*, Hs.612374, *GUCY1B3***

Data refer to the following comparisons: DS AMKL (pool 'A') vs. MK (pool 'D'). Analysis was performed using default parameters (see Methods section). Segments are sorted by decreasing 'A'/'D' ratio. In the "Map" mode, TRAM displays UniGene EST clusters (with the prefix "Hs." in the case of *Homo sapiens*) only if they have an expression value. Some segments are not shown for simplicity because they are over-lapping with those highlighted in one of the listed regions. The complete results for these models are available as on line additional material. [a] Chr: chromosome. The segment location cytoband was derived from that of the first mapped gene within the segment. [b] Segment Start/End: chromosomal coordinates for each segment. [c] Significantly over-/under-expressed genes as marked in the TRAM results.

At single gene level, a fold increase higher than 18 was observed in all of the first 20 genes with the greatest expression ratios of DS AMKL vs. MK samples (Table 22 and Additional file

4, the last is available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/). In particular, a 59-fold increase was observed for *TMEM241*, encoding a transmembrane protein of unknown function. Among the genes with the lowest 'A'/'D' expression ratio a 589.7-fold decrease was observed for *TSPAN10*, encoding for tetraspanin 10.

Table 22 List of the five most over- or under-expressed genes (all significantly, with q<0.05) in the DS AMKL vs. MK transcriptome map.

Gene	Value 'A'	Value 'D'	'A'/'D' Ratio	Location	Data points		SD	SD
					'A'	'D'	'A'	'D'
Over-expressed genes								
TMEM241	1,036.6	17.6	58.885	18q11.2	18	8	57.7	115.8
CMBL	592.4	11.7	50.759	5p15.2	27	8	142.0	30.1
IFI27	467.8	11.6	40.474	14q32	36	82	90.0	27.6
CSRNP1	640.6	16.8	38.023	3p22	9	8	41.3	138.3
PTGS2	2,205.0	59.2	37.276	1q25.2 q25.3	63	20	156.4	131.6
APOC2[ab]	431.7	11.7	37.044	19q13.2	54	28	81.4	41.6
SLITRK6[b]	465.5	13.3	35.057	13q31.1	27	10	113.0	95.6
Under-expressed genes								
MAP3K10	5.4	2,352.3	0.002	19q13.2	36	19	47.4	135.1
DRD4	5.0	2,327.6	0.002	11p15.5	36	19	51.3	135.2
DLGAP3	4.8	2,297.9	0.002	1p35.3-p34.1	9	7	40.8	37.6
LBX1	2.7	1,358.5	0.002	10q24	36	19	46.4	152.6
TSPAN10[a]	7.9	4,674.7	0.002	17q25.3	9	8	55.0	46.0

Data refer to the following comparisons: DS AMKL (pool 'A') vs. MK (pool 'D'). Value: mean gene expression value normalized across all the pool samples; Data points: number of spots associated to an expression value for the locus; SD: standard deviation for the expression value expressed as percentage of the mean. Full results available as additional material (see text). [a] The segment window contains more than one gene, but the significance is assumed to be maintained because the expression value of this over- or under-expressed gene prevails over the others.
[b] This gene, exceeding the limit of five genes for each list, has been shown for its relevance in the Discussion, being recurrent in other comparisons.

At chromosomal level, the highest ratio was observed for chr21 (1.75), the lowest for chrX (1.17), other values were in the range from 1.68 (chrY) to a value of 1.23 for chr17.

Regarding the non-DS AMKL samples, results obtained by default analysis and derived from 'B'/'D' expression ratio for 25,819 loci included one significantly over- and one significantly under-expressed segment (Table 23). The highest expression ratio (22.30) between non-DS AMKL and normal MK was observed in the same segment on chr15, significantly over-expressed also in the DS AMKL transcriptome map. This segment was the only significantly over-expressed one in this comparison. Similarly, the only significantly under-expressed segment includes the cluster of MK specific genes on chromosome 4 also found to be under-expressed in DS AMKL samples (Table 21).

Table 23 Genomic segments significantly over- or under-expressed in the non-DS AMKL vs. MK differential transcriptome map.

Chr and Location[a]	Segment Start[b]	Segment End[b]	'B'/'D' Ratio	q-value	Genes in the segment[c]
Over-expressed segments					
chr15 15q26.2	96,750,001	97,250,000	22.30	<0.00001	**Hs.677040, Hs.661950, *NR2F2-AS1*, Hs.592015**
Under-expressed segments					
chr4 4q13-q21	74,500,001	75,000,000	0.39	0.00009	***PF4V1*, *PF4*, *PPBP*, *CXCL3***

Data refer to the following comparisons: non-DS AMKL (pool 'B') vs. MK (pool 'D'). Analysis was performed using default parameters (see Methods section). Segments are sorted by decreasing 'B'/'D' ratio. In the "Map" mode, TRAM displays UniGene EST clusters (with the prefix "Hs." in the case of *Homo sapiens*) only if they have an expression value. Some segments are not shown for simplicity because they are over-lapping with those highlighted in one of the listed regions. The complete results for these models are available as on line additional material. [a] Chr: chromosome. The segment location cytoband was derived from that of the first mapped gene within the segment. [b] Segment Start/End: chromosomal coordinates for each segment. [c] Significantly over-/under-expressed genes as marked in the TRAM results.

At single gene level, a fold increase higher than 31 was observed in all of the first 20 genes with the greatest expression ratios of non-DS AMKL vs. MK samples (Table 24 and Additional file 5, the last is available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/). In particular, a 156-fold increase was observed for *TMEM241*. Overall, there was a remarkable overlap between the most over- (*TMEM241*, *CMBL*, *ZNF445*, *SPRR4*) and under-expressed (*PF4V1*, *FLJ22184*, *FSIP2*, *PPP1R3B*, *HIST3H3*, *PIF1*, *SPSB4*, *ILDR1*, *MAP3K10*, *DRD4*, *LBX1*, *TSPAN10*) genes in DS and in non-DS AMKL samples (Additional files 4 and 6, available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/).

Table 24 List of the five most over- or under-expressed genes (all significantly, with q<0.05) in the non-DS AMKL vs. MK transcriptome map.

Gene	Value 'B'	Value 'D'	'B'/'D' Ratio	Location	Data points		SD	SD
					'B'	'D'	'B'	'D'
Over-expressed genes								
TMEM241	2,754.1	17.6	156.443	18q11.2	14	8	89.9	115.8
CDYL2	1,613.5	13.1	122.994	16q23.2	7	8	128.9	69.8
MPV17L	1,857.8	24.4	76.280	16p13.11	7	8	119.5	65.0
ATP6V0D2	836.7	11.9	70.488	8q21.3[b]	21	10	169.4	82.9
CMBL	757.9	11.7	64.938	5p15.2	21	8	176.2	30.1
Under-expressed genes								
ILDR1	13.7	4,333.8	0.003	3q13.33	14	9	98.4	63.3
HIST3H3	3.9	1,313.0	0.003	1q42	40	19	78.0	137.9
TSPAN10[a]	12.1	4,674.7	0.003	17q25.3	7	8	88.1	46.0
MAP3K10	3.9	2,352.3	0.002	19q13.2	45	19	40.1	135.1
LBX1	2.0	1,358.5	0.001	10q24	45	19	85.7	152.6

Data refer to the following comparisons: non-DS AMKL (pool 'B') vs. MK (pool 'D'). Value: mean gene expression value normalized across all the pool samples; Data points: number of spots associated to an expression value for the locus; SD: standard deviation for the expression value expressed as percentage of the mean. Full results available as additional material (see text). [a] The segment window contains more than one gene, but the significance is assumed to be maintained because the expression value of this over- or under-expressed gene prevails over the others. [b] Cytoband not available in Gene was derived from the UCSC Genome Browser (University of California Santa Cruz (UCSC) Genome Browser. (http://genome-euro.ucsc.edu/cgi-bin/hgGateway).

At chromosomal level, looking at the median 'B'/'D' expression ratio for the genes located in the same chromosome, the highest ratio was observed for chr21 (1.77) and chrY (1.73), followed by chr13 (1.71) and chr10 (1.59); other values were in the range from 1.58 (chr20), with a value of 1.24 for chrX.

4.2.7 Transcriptome map comparison of DS AMKL or non DS-AMKL vs. normal CB MK

Regional differential expression of pool 'A' (43 DS AMKL samples) or pool 'B' (45 non-DS AMKL samples) versus pool 'E' (8 normal cord-blood (CB)-derived MK cell samples, 191,798 data points) (Table 18) was investigated. The main results are shown in Figure 9.

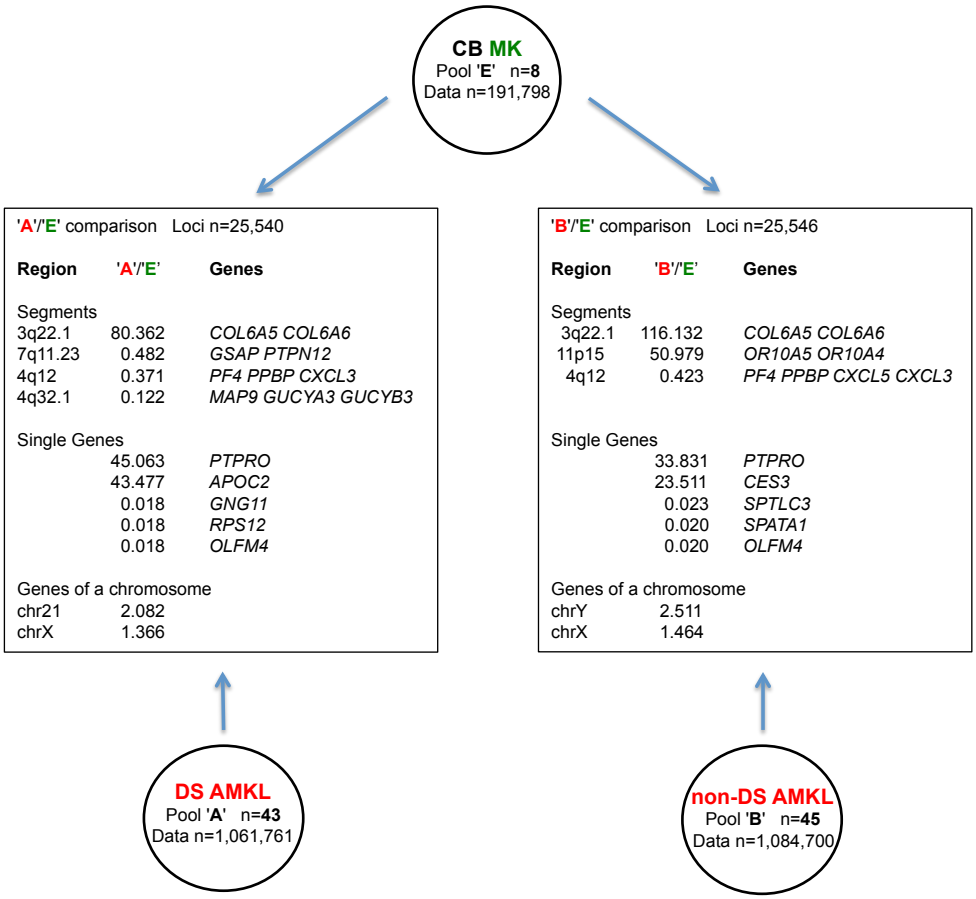


Figure 9 Main results of DS AMKL vs. CB MK and non-DS AMKL vs. CB MK comparisons. For each comparison the number of loci analyzed, the most over- or under-expressed segments and single genes, and the highest and lowest median expression ratios for all the genes located in the same chromosome are indicated.

For what DS AMKL samples are concerned, results derived from 'A'/'E' expression ratio for 25,540 loci (Additional file 6, available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/) included 1 significantly over- and 3 under-expressed segments in DS AMKL cells (Table 25). A remarkable expression ratio (80.36) between DS AMKL cells and normal CB MK was observed for a segment on chromosome 3 (3q22.1), including collagen-encoding *COL6A5* and *COL6A6* known loci. The three significantly under-expressed segments included the region on chromosome 4 (4q12-q21) with *PF4*, *PPBP* and *CXCL3* loci implied in MK differentiation.

Table 25. Genomic segments significantly over- or under-expressed in the DS AMKL vs. normal CB MK differential transcriptome map.

Chr and Location[a]	Segment Start[b]	Segment End[b]	'A'/'E' Ratio	q-value	Genes in the segment[c]
Over-expressed segments					
chr3 3q22.1	130,000,001	130,500,000	80.36	0.00015	***COL6A5*, *COL6A6*, Hs.596805**
Under-expressed segments					
chr7 7q11.23	77,000,001	77,500,000	0.48	0.00117	***GSAP*, *PTPN12*, Hs.720279**
chr4 4q12-q21	74,750,001	75,250,000	0.37	0.00119	***PF4*, *PPBP*, *CXCL3***
chr4 4q32.1	156,250,001	156,750,000	0.12	0.00000	***MAP9*, *GUCY1A3*, Hs.612374, *GUCY1B3***

Data refer to the following comparisons: DS AMKL (pool 'A') vs. normal CB MK (pool 'E'). Analysis was performed using default parameters (see Methods section). Segments are sorted by decreasing 'A'/'E' ratio. In the "Map" mode, TRAM displays UniGene EST clusters (with the prefix "Hs." in the case of *Homo sapiens*) only if they have an expression value. Some segments are not shown for simplicity because they are over-lapping with those highlighted in one of the listed regions. The complete results for these models are available as on line additional material. [a] Chr: chromosome. The segment location cytoband was derived from that of the first mapped gene within the segment. [b] Segment Start/End: chromosomal coordinates for each segment. [c] Significantly over-/under-expressed genes as marked in the TRAM results.

At single gene level, a fold increase higher than 15.7 was observed in all of the first 20 genes with the greatest expression ratios of DS AMKL cells vs. CB MK (Table 26 and Additional file 6, the last is available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/ - where Pool 'D' here is named 'A' and Pool 'H' here is named 'B'). In particular, a 45-fold increase was observed for the tyrosine phosphatase receptor gene (*PTPRO*), known to be involved in megakaryocytopoiesis.

Table 26. List of the five most over- or under-expressed genes (all significantly, with q<0.05) in the DS AMKL vs. normal CB MK differential transcriptome map.

Gene	Value 'A'	Value 'E'	Ratio 'A'/'E'	Location	Data Points		SD	SD
					'A'	'E'	'A'	'E'
Over-expressed genes								
PTPRO	679.7	15.1	45.063	12p13-p12	81	9	133.0	60.8
APOC2[a]	431.7	9.9	43.477	19q13.2	54	10	81.4	43.4
IFI27	467.8	11.5	40.742	14q32	36	8	90.0	57.1
HDC	1,119.4	34.2	32.716	15q21-q22	36	8	98.8	168.1
VIPR2[a b]	346.2	13.5	25.582	7q36.3	115	23	334.8	49.3
Under-expressed genes								
CDKL1	4.6	174.3	0.027	14q21.3	45	9	65.3	99.0
ASAP2	19.1	734.1	0.026	2p24	36	8	95.4	67.3
GNG11	18.9	1,042.3	0.018	7q21	36	8	119.7	81.2
RPS12	20.7	1,157.2	0.018	6q23.2	43	15	113.7	112.4
OLFM4	8.9	499.3	0.018	13q14.3	36	7	83.9	121.5

Data refer to the following comparisons: DS AMKL (pool 'A') vs. normal CB MK (pool 'E'). Value: mean gene expression value normalized across all the pool samples; Data points: number of spots associated to an expression value for the locus; SD: standard deviation for the expression value expressed as percentage of the mean. Full results available as additional material (see text). [a] The segment window contains more than one gene, but the significance is assumed to be maintained because the expression value of this over- or under-expressed gene prevails over the others. [b] This gene is one of the five most over-expressed ones, but the value is not statistically significant due to the presence of a lot of loci associated with less than 5 data points in the CB MK integrated dataset.

At chromosomal level, the highest ratio was observed for chr21 (2.08), followed by chrY (1.89); other values were in the range from 1.84 (chr22) to 1.37 (chrX).

Regarding the non-DS AMKL samples, results derived from 'B'/'E' expression ratio (for 25,546 loci) included 2 significantly over- and 1 under-expressed segments in non-DS AMKL (Table 27). A remarkable expression ratio between non-DS AMKL and normal CB MK (116.13) was observed in the same segment on chromosome 3 (3q22.1), including collagen-encoding *COL6A5* and *COL6A6* known loci that was observed in DS AMKL samples. The second segment was specific of non-DS AMKL samples and included the two olfactive receptor genes *OR10A5* and *OR10A4*. The only significantly under-expressed segment included the region on chromosome 4 (4q12-q21) highly enriched in MK-specific loci (*PF4*, *PPBP*, *CXCL5* and *CXCL3*) as in the case of DS AMKL samples, and was extended to *PF4V1* locus.

Table 27. Genomic segments significantly over- or under-expressed in the non-DS AMKL vs. normal CB MK differential transcriptome map.

Chr and Location[a]	Segment Start[b]	Segment End[b]	'B'/'E' Ratio	q-value	Genes in the segment[c]
Over-expressed segments					
chr3 3q22.1	130,000,001	130,500,000	116.13	0.00030	***COL6A5*, *COL6A6*, Hs.596805**
chr11 11p15	6,750,001	7,250,000	50.98	0.00117	***OR10A5*, *OR10A4*, *LOC100506238***
Under-expressed segments					
chr4 4q12-q21	74,750,001	75,250,000	0.42	0.00003	***PF4*, *PPBP*, *CXCL5*, *CXCL3***

Data refer to the following comparisons: non-DS AMKL (pool 'B') vs. normal CB MK (pool 'E'). Analysis was performed using default parameters (see Methods section). Segments are sorted by decreasing 'B'/'E' ratio. In the "Map" mode, TRAM displays UniGene EST clusters (with the prefix "Hs." in the case of *H. sapiens*) only if they have an expression value. Some segments are not shown for simplicity because they are over-lapping with those highlighted in one of the listed regions. The complete results for this model are available as on line additional material. [a] Chr: chromosome. The segment location cytoband was derived from that of the first mapped gene within the segment. [b] Segment Start/End: chromosomal coordinates for each segment. [c] Significantly over-/under-expressed genes as marked in the TRAM results.

At single gene level, a fold increase higher than 14.7 was observed in all of the first 20 genes with the greatest expression ratios of non-DS AMKL vs. CB MK samples (Table 28 and Additional file 7, the last is available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/). In particular, a 33.8-fold increase was observed for *PTPRO*, encoding a tyrosine phosphatase receptor. Overall, there was some overlapping between the most over- and under-expressed genes in DS and in non-DS AMKL samples (Additional files 6 and 7, available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/).

Table 28 List of the five most over- or under-expressed genes (all significantly, with q<0.05) in the non-DS AMKL vs. normal CB MK differential transcriptome map.

Gene	Value 'B'	Value 'E'	Ratio 'B'/'E'	Location	Data Points		SD	SD
					'B'	'E'	'B'	'E'
Over-expressed genes								
PTPRO[b]	510.3	15.1	33.831	12p13-p12	97	9	137.8	60.8
CES3[b]	287.4	12.2	23.511	16q22.1	47	9	335.8	73.8
SCD5[a b]	544.7	27.2	19.992	4q21.22	61	18	290.3	73.9
TOP3B[a b]	150.2	7.5	19.922	22q11.22	60	14	431.8	19.7
MT1E[a b]	237.1	11.9	19.899	16q13	45	8	92.5	27.3
Under-expressed genes								
Hs.23729[a]	2.3	85.1	0.027	1p13.2[c]	40	8	121.0	89.5
PDE6C	2.2	88.2	0.025	10q24	45	8	112.9	105.9
SPTLC3	4.8	203.7	0.023	20p12.1	47	9	86.5	110.7
SPATA1	7.0	356.2	0.020	1p22.3	40	7	127.4	77.9
OLFM4	9.7	499.3	0.020	13q14.3	45	7	105.1	121.5

Data refer to the following comparisons: non-DS AMKL (pool 'B') vs. normal CB MK (pool 'E'). Value: mean gene expression value normalized across all the pool samples; Data points: number of spots associated to an expression value for the locus; SD: standard deviation for the expression value expressed as percentage of the mean. Full results available as additional material (see text). [a] The segment window contains more than one gene, but the significance is assumed to be maintained because the expression value of this over- or under-expressed gene prevails over the others. [b] This gene is one of the five most over-expressed ones, but the value is not statistically significant due to the presence of a lot of loci associated with less than 5 data points in the CB MK integrated dataset. [c] Cytoband not available in Gene was derived from the UCSC Genome Browser (University of California Santa Cruz (UCSC) Genome Browser. (http://genome-euro.ucsc.edu/cgi-bin/hgGateway).

At chromosomal level, regarding the median 'B'/'E' expression ratio for the genes located in the same chromosome, the highest ratio was observed for chrY (2.51), followed by chr21 (2.19); other values were in the range from 2.03 (chr20) to 1.46 (chrX).

4.2.8 Transcriptome map comparison of DS AMKL vs. TMD

Regional differential expression of pool 'A' (43 DS AMKL samples) versus pool 'C' (20 TMD samples, 398,162 data points) (Table 18) was investigated. The main results are shown in Figure 10.

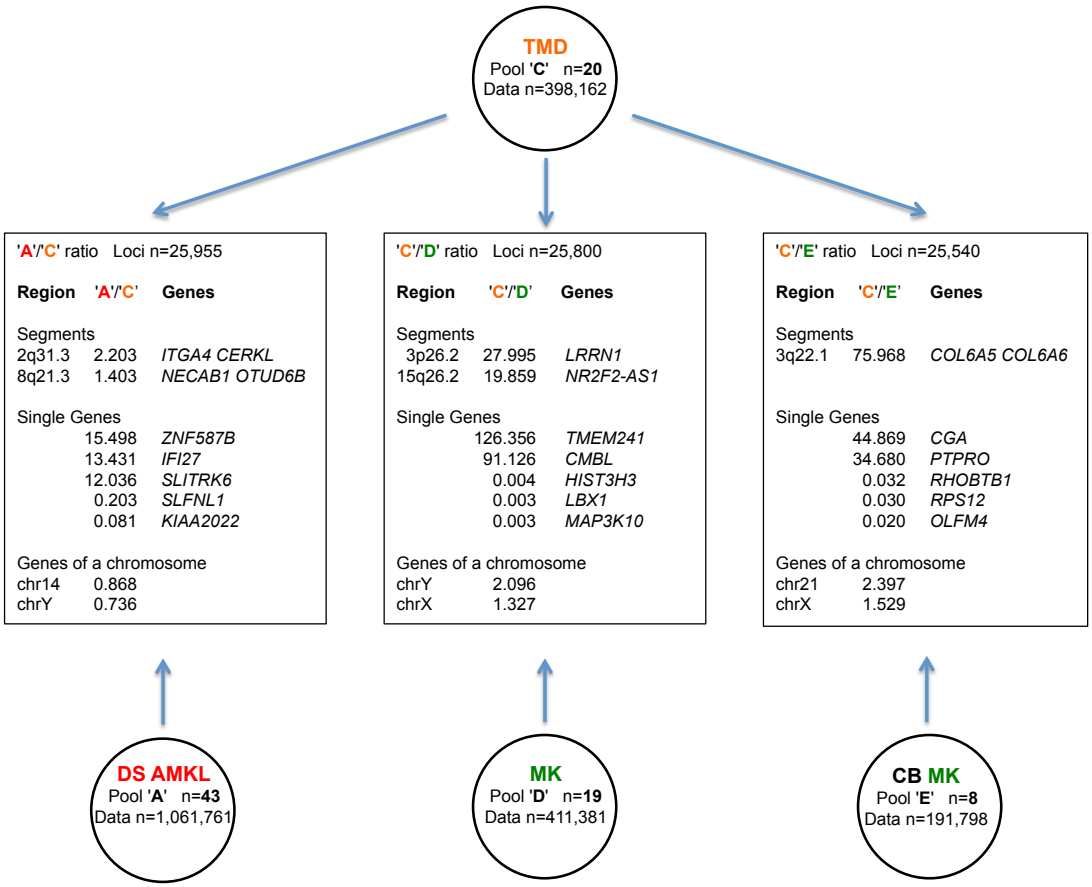


Figure 10 Main results of DS AMKL vs. TMD, TMD vs. MK, TMD vs. CB MK comparisons. For each comparison the number of loci analyzed, the most over- or under-expressed segments and single genes, and the highest and lowest median expression ratios for all the genes located in the same chromosome are indicated.

Results obtained by default analysis and derived from 'A'/'C' expression ratio for 25,955 loci (Additional file 8, it is available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/) included 2 significantly over-expressed segments in DS AMKL (Table 29). The highest expression ratio (2.20) between DS AMKL and TMD was observed in a segment on chromosome 2 (2q31.3), including the known gene *ITGA4*, encoding an alpha 4 chain of integrin protein and *CERKL*, a gene responsible for retinitis pigmentosa and involved in the protection of cells from apoptosis induced by oxidative stress (Repeat Masker (http://www.repeatmasker.org/). The second segment with the highest expression ratio (1.40) was located on chromosome 8 (8q21.3) and contained the *NECAB1* and *OTUD6B* genes, encoding for neuronal Ca(2+)-binding protein and the deubiquitinating enzyme, respectively.

Table 29 Genomic segments significantly over- or under-expressed in the DS AMKL vs. TMD differential transcriptome map.

Chr[a] and Location	Segment Start[b]	Segment End[b]	'A'/'C' Ratio	q-value	Genes in the segment[c]
Over-expressed segments					
chr2 2q31.3	182,000,001	182,500,000	2.20	0.00045	***ITGA4*, *CERKL*, Hs.72981**
chr8 8q21.3	91,750,001	92,250,000	1.40	0.00051	***NECAB1*, *LOC100127983*, *OTUD6B***

Data refer to the following comparisons: DS AMKL (pool 'A') vs. TMD (pool 'C'). Segments are sorted by increasing 'A'/'C'. In the "Map" mode, TRAM displays UniGene EST clusters (with the prefix "Hs." in the case of *H. sapiens*) only if they have an expression value. Some segments are not shown for simplicity because they are over-lapping with those highlighted in one of the listed regions. The complete results for this model are available as on line additional material. [a] Chr: chromosome. The segment location cytoband was derived from that of the first mapped gene within the segment. [b] Segment Start/End: chromosomal coordinates for each segment. [c] Significantly over-/under-expressed genes as marked in the TRAM results.

At single gene level, a fold increase ranged from 15.5 to 3.5 for the first 20 genes with the greatest expression ratios of DS AMKL vs. TMD samples (Table 30 and Additional file 8, the last is available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/). The highest fold increases were observed for the *ZNF587B* (15.5) and *IFI27* (13.4) genes, encoding for a zinc finger protein and the interferon alpha-inducible protein 27, respectively. The lowest 'A'/'C' expression ratios were observed for *KIAA2022* (10-fold decrease) and *SLFNL1* (5- fold decrease) genes.

Table 30 List of the five genes most over- or under-expressed (all significantly, with q<0.05) in the DS AMKL vs. TMD differential transcriptome map.

Gene	Value 'A'	Value 'C'	Ratio 'A'/'C'	Location	Data Points		SD	SD
					'A'	'C'	'A'	'C'
Over-expressed genes								
ZNF587B	615.0	39.7	15.498	19q13.43[b]	43	20	427.6	20.9
IFI27	467.8	34.8	13.431	14q32	36	11	90.0	91.4
SLITRK6	465.5	38.7	12.036	13q31.1	27	9	113.0	84.5
BST2[a]	117.6	14.8	7.967	19p13.1	36	11	260.0	98.8
ZNF521	143.0	19.6	7.309	18q11.2	18	6	113.5	58.6
Under-expressed genes								
ALS2CR12	4.2	15.8	0.268	2q33.1	18	6	80.3	103.2
GBP5	11.6	43.8	0.266	1p22.2	18	6	75.7	110.6
GHRH	14.7	67.1	0.219	20q11.2	43	20	97.7	257.0
SLFNL1[a]	18.9	92.9	0.203	1p34.2	18	6	90.3	122.1
KIAA2022	24.3	297.9	0.081	Xq13.3	27	9	52.2	260.9

Data refer to the following comparisons: DS AMKL (pool 'A') vs. TMD (pool 'C'). Value: mean gene expression value normalized across all the pool samples; Data points: number of spots associated to an expression value for the locus; SD: standard deviation for the expression value expressed as percentage of the mean. Full results available as additional material (see text).[a]The segment window contains more than one gene, but the significance is assumed to be maintained because the expression value of this over- or under-expressed gene prevails over the others. [b]Cytoband not available in Gene was derived from the UCSC Genome Browser (University of California Santa Cruz (UCSC) Genome Browser. (http://genome-euro.ucsc.edu/cgi-bin/hgGateway).

At chromosomal level, the highest ratio was observed for chr14 (0.87, followed by 0.86 for chr5, chr21, chr8, chr12 and chr16), the lowest for chrY (0.74).

4.2.9 Transcriptome map comparison of TMD vs. normal MK or CB MK cells

Regional differential expression of pool 'C' (20 TMD samples) versus pool 'D' (19 MK samples) or 'E' (8 CB MK samples) (Table 18) was investigated. The main results are shown in Figure 10.

For what MK samples are concerned, results obtained by default analysis and derived from 'C'/'D' expression ratios for 25,800 loci (Additional file 9, available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/) included 2 significantly over-expressed segments in TMD cells (Table 31).

Table 31 Genomic segments significantly over- or under-expressed in the TMD vs. normal MK differential transcriptome map.

Chr and Location[a]	Segment Start[b]	Segment End[b]	'C'/'D' Ratio	q-value	Genes in the segment[c]
Over-expressed segments					
chr3 3p26.2	3,500,001	4,000,000	28.00	0.00006	**Hs.241414,**[d] ***LRRN1*****, Hs.128128**
chr15 15q26.2	96,750,001	97,250,000	19.86	0.00000	**Hs.677040,**[d] **Hs.661950,**[d] ***NR2F2-AS1*****, Hs.592015**

Data refer to the following comparisons: TMD (pool 'C') vs. normal MK (pool 'D'). Analysis was performed using default parameters (see Methods section). Segments are sorted by increasing 'C'/'D' ratio. In the "Map" mode, TRAM displays UniGene EST clusters (with the prefix "Hs." in the case of *H. sapiens*) only if they have an expression value. Some segments are not shown for simplicity because they are over-lapping with those highlighted in one of the listed regions. The complete results for this model are available as on line additional material.
[a] Chr: chromosome. The segment location cytoband was derived from that of the first mapped gene within the segment. [b] Segment Start/End: chromosomal coordinates for each segment. [c] [c] Significantly over-/under-expressed genes as marked in the TRAM results. [d] This UniGene cluster contains EST with at least one Alu sequence, according to Repeat Masker (http://www.repeatmasker.org/). For this reason, it can not be excluded that its over-expression is related to unspecific hybridization by Alu-containg probes.

The highest expression ratio (28.0) between TMD and normal MK was observed in a segment on chromosome 3 including the known gene *LRRN1*, already observed as over-expressed in comparison of DS AMKL vs. normal MK (Table 21). The second segment with the highest expression ratio (19.9) was located on chromosome 15, and contained the locus *NR2F2-AS1*, encoding for an antisense mRNA, already observed as over-expressed in comparison of DS AMKL and non-DS AMKL vs. normal MK (Table 21 and 23).

At single gene level, fold-increase was higher than 16.5 for the first 20 genes with the greatest expression ratios (Table 32 and Additional file 9, the last is available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/), with the highest fold increases for *TMEM241*

(126.4) and the cysteine hydrolase gene (*CMBL*) (91.1). The lowest 'C'/'D' expression ratios were observed for a member of the serine/threonine kinase family (*MAP3K10*) and the homeobox gene (*LBX1*) (both with 333-fold decrease) .

At chromosomal level, the highest ratio was observed for chrY (2.1) and chr21 (1.93), followed by chr13 (1.71), the lowest for chrX (1.33).

Table 32 List of the five genes most over- or under-expressed (all significantly, with q<0.05) in the TMD vs. normal MK differential transcriptome map.

Gene	Value 'C'	Value 'D'	Ratio 'C'/'D'	Location	Data Points		SD	SD
					'C'	'D'	'C'	'D'
Over-expressed genes								
TMEM241	2,224.4	17.6	126.356	18q11.2	6	8	57.1	115.8
CMBL	1,063.5	11.7	91.126	5p15.2	9	8	168.8	30.1
PTGS2	2,256.4	59.2	38.145	1q25.2-q25.3	20	20	147.6	131.6
CGA	360.6	10.0	36.238	6q12-q21	14	19	159.7	27.3
TMEFF2	755.1	22.5	33.605	2q32.3	9	9	190.3	44.3
Under-expressed genes								
ILDR1	18.3	4,333.8	0.004	3q13.33	6	9	58.9	63.3
DRD4	9.3	2,327.6	0.004	11p15.5	11	19	52.3	135.2
HIST3H3	5.2	1,313.0	0.004	1q42	11	19	36.2	137.9
LBX1	4.3	1,358.5	0.003	10q24	11	19	26.6	152.6
MAP3K10	7.4	2,352.3	0.003	19q13.2	11	19	28.8	135.1

Data refer to the following comparisons: TMD (pool 'C') vs. normal MK (pool 'D'). Value: mean gene expression value normalized across all the pool samples; Data points: number of spots associated to an expression value for the locus; SD: standard deviation for the expression value expressed as percentage of the mean. Full results available as additional material (see text).

As far as the comparison of TMD with CB MK samples is concerned, results derived from the 'C'/'E' expression ratio for 25,540 loci (Additional file 10, available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/) included only 1 significantly over-expressed segment in TMD cells (Table 33). The segment with a significant high expression ratio (76.0) between TMD and normal CB MK cells was on chromosome 3 (3q22.1), including the known genes *COL6A5* and *COL6A6* and already observed as over-expressed in DS as well in non-DS AMKL samples in comparison with CB MK samples.

Table 33 Genomic segments significantly over- or under-expressed in the TMD vs. normal CB MK differential transcriptome map.

Chr and Location[a]	Segment Start[b]	Segment End[b]	'C'/'E' Ratio	q-value	Genes in the segment[c]
chr3 3q22.1	130,000,001	130,500,000	75.97	0.00015	***COL6A5, COL6A6*, Hs.596805**

Data refer to the following comparisons: TMD (pool 'C') vs. normal CB MK (pool 'E'). Analysis was performed using default parameters (see Methods section). Segments are sorted by 'C'/'E' ratio. In the "Map" mode, TRAM displays UniGene EST clusters (with the prefix "Hs." in the case of *H. sapiens*) only if they have an expression value. Some segments are not shown for simplicity because they are over-lapping with those highlighted in one of the listed regions. The complete results for this model are available as additional material on line. [a] Chr: chromosome. The segment location cytoband was derived from that of the first mapped gene within the segment.[b] Segment Start/End: chromosomal coordinates for each segment. [c] Significantly over-/under-expressed genes as marked in the TRAM results.

At single gene level, a fold increase ranged from 44.9 to 15.1 for the first 20 genes with the greatest expression ratios (Table 34 and Additional file 10, the last is available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/). The highest fold increases were observed for *CGA* (44.9), encoding for the alpha chain of the glycoprotein hormones and *PTPRO* (34.7), as already observed in non-DS AMKL vs. CB MK comparison. The lowest 'C'/'E' expression ratio was observed for *OLFM4* (1:50 fold decrease), encoding olfactomedin 4, an antiapoptotic factor that promotes tumor growth.

At chromosomal level, the highest ratio was observed for chr21 (2.40), followed by chrY (2.37) and chr20 (2.13), the lowest for chrX (1.53).

Table 34 List of the five genes most over- or under-expressed (all significantly, with q<0.05) in the TMD vs. normal CB MK differential transcriptome map.

Gene	Value 'C'	Value 'E'	Ratio 'C'/'E'	Location	Data Points		SD	SD
					'C'	'E'	'C'	'E'
Over-expressed genes								
CGA	360.6	8.0	44.869	6q12-q21	14	8	159.7	39.9
PTPRO[b]	523.1	15.1	34.680	12p13-p12	25	9	116.8	60.8
CLCA1[b]	319.2	9.9	32.380	1p22.3	20	8	116.7	17.5
APOC2[a b]	301.6	9.9	30.372	19q13.2	17	10	66.6	43.4
HDC [a b]	918.0	34.2	26.830	15q21-q22	11	8	59.9	168.1
Under-expressed genes								
CDKL1	5.9	174.3	0.034	14q21.3	14	9	82.0	99.0
GNG11	34.3	1,042.3	0.033	7q21	11	8	80.1	81.2
RHOBTB1	7.7	241.3	0.032	10q21.2	20	8	49.7	97.4
RPS12	34.6	1,157.2	0.030	6q23.2	20	15	77.4	112.4
OLFM4	10.0	499.3	0.020	13q14.3	11	7	56.0	121.5

Data refer to the following comparisons: TMD (pool 'C') vs. normal CB MK (pool 'E'). Value: mean gene expression value normalized across all the pool samples; Data points: number of spots associated to an expression value for the locus; SD: standard deviation for the expression value expressed as percentage of the mean. Full results available as additional material (see text).[a] The segment window contains more than one gene, but the significance is assumed to be maintained because the expression value of this over- or under-expressed gene prevails over the others. [b] This gene is one of the five most over-expressed ones, but the value is not statistically significant due to the presence of a lot of loci associated with less than 5 data points in the CB MK integrated dataset.

4.2.10 Comparison with previously published data

As a result of the analysis above described, a reference integrated map for the expression of about 26,000 mapped sequences (~75% known genes and ~25% expression sequence tags - ESTs) was de facto generated for five cell types (DS AMKL cells, non-DS AMKL cells, TMD cells, MK and CB MK). This gave us the opportunity to compare our data with the expression values of specific known genes from previously published works about the considered cell types.

Following analysis of the main literature about AMKL, we selected 38 genes of interest and have tabulated their expression values desumed from our 9 differential maps, comparing these values to the ones previously described in different experimental settings (Table 35).

The wide agreement of expression ratio values for specific genes between our data, generated by systematic meta-analysis of hundred of thousands of gene expression values from any gene expression profile available, and the data obtained by different marker-specific methods in published quantitative studies, is relevant for the validation of our maps that may so be used for exploring any other expression ratio in the considered biological conditions.

Table 35 Comparison with previously published data.

Gene	Location	References	'A'/'B'	'A'/'D'	'B'/'D'	'A'/'E'	'B'/'E'	'A'/'C'	'C'/'D'	'C'/'E'
MK Markers										
BST2	19p13.11	Ge 2006	**0.53**	5.36	10.07	7.21	13.56	7.97	0.67	0.91
GATA1	Xp11.23	Bourquin 2006 Khan 2011 Melemed 1997	**3.39**	**1.45**	**0.43**	1.97	0.58	**1.19**	**1.22**	**1.66**
GP1BA	17p13.2	Fuhrken 2007	0.46	**0.07**	**0.14**	**0.25**	**0.53**	0.81	0.08	0.30
HBG1	11p15.5	Bourquin 2006	**2.36**	1.40	0.59	1.07	0.45	0.78	1.79	1.37
HBG2	11p15.5	Bourquin 2006	**2.47**	1.31	0.53	1.06	0.43	0.71	1.84	1.48
ITGA2B	17q21.32	Ge 2006	0.82	**0.63**	0.77	1.52	1.84	1.05	0.60	1.44
ITGB3	17q21.32	Ge 2006	0.75	**0.08**	**0.10**	**0.10**	**0.13**	0.65	0.12	0.15
PF4	4q12-q21	Lenzi 2011	0.90	**0.06**	**0.06**	**0.11**	**0.12**	0.44	0.13	0.25
PPBP	4q12-q13	Lenzi 2011	1.45	**0.10**	**0.07**	**0.13**	**0.09**	0.28	0.34	0.44
TAL1	1p32	Ge 2006	**0.85**	**0.19**	**0.22**	**0.16**	**0.19**	1.05	0.18	0.16
Chr21 Genes										
ADAMTS1	21q21.2	Strippoli 2013	6.19	**7.02**	1.14	**7.14**	1.15	1.82	**3.85**	**3.91**
BACH1	21q22.11	Bourquin 2006	**1.86**	1.44	0.77	1.04	0.56	1.55	0.93	0.67
DYRK1A	21q22.13	Herrera 1998	**1.61**	0.87	0.54	0.63	0.39	**1.14**	0.77	0.55
ERG	21q22.3	Bourquin 2006 Klusmann 2010b	**0.68**	3.47	5.12	3.10	4.57	1.42	2.44	2.18
ETS2	21q22.2	Bourquin 2006 Klusmann 2010b	**0.93**	1.11	1.20	1.39	1.49	0.73	1.52	1.90
GABPA	21q21.3	Bourquin 2006 Klusmann 2010b	**1.39**	2.45	1.76	1.52	1.10	1.55	1.58	0.98
RUNX1	21q22.3	Bourquin 2006 Klusmann 2010b	**0.71**	1.08	1.51	0.84	1.18	1.06	1.01	0.80
SOD1	21q22.11	Tonelli 2000	**1.88**	1.52	0.81	2.78	1.48	1.52	1.00	1.84
SON	21q22.11	Bourquin 2006	**1.28**	1.39	1.08	1.03	0.80	1.44	0.96	0.71
Other genes										
APOC1	19q13.2	Bourquin 2006	**4.37**	10.04	2.30	16.26	3.72	**1.79**	5.63	9.11

		Ge 2006 Lightfoot 2004								
APOC2	19q13.2	Bourquin 2006	**4.36**	37.04	8.49	43.48	9.96	1.43	25.88	30.37
APOE	19q13.2	Bourquin 2006 Ge 2006	**2.18**	3.57	1.64	3.24	1.48	1.56	2.30	2.08
CDA	1p36.2-p35	Ge 2006	**0.39**	0.40	1.03	0.88	2.24	0.88	0.46	1.00
DICER1	14q32.13	Klusmann 2010b	0.80	0.41	0.51	0.27	0.34	0.99	0.41	0.27
FCER1A	1q23	Bourquin 2006 Ge 2006	**6.32**	4.28	0.68	8.86	1.40	1.32	3.25	6.72
GYPA	4q31.21	Ge 2006	**3.40**	1.56	0.46	0.89	0.26	1.19	1.31	0.75
GYPB	4q31.21	Ge 2006	**2.81**	1.28	0.46	0.89	0.31	1.27	1.01	0.70
GYPE	4q31.1	Ge 2006	**2.07**	1.63	0.79	1.71	0.82	1.72	0.95	0.99
HDC	15q21-q22	Bourquin 2006 Ge 2006	**17.10**	9.95	0.58	32.72	1.91	1.22	8.16	26.83
IGF1R	15q26.3	Klusmann 2010a	0.76	**1.18**	**1.55**	0.94	1.24	1.21	0.98	0.78
ITGAL	16p11.2	Taniguchi 1999	0.79	**0.21**	**0.26**	**0.09**	**0.11**	**1.02**	**0.20**	**0.09**
KIT	4q11-q12	Bourquin 2006	**1.66**	3.34	2.01	3.08	1.85	1.22	2.74	2.53
MPL	1p34	Gear and Camerini 2003	1.11	**0.31**	**0.28**	0.22	0.20	1.23	**0.25**	0.18
MTOR	1p36.22	Klusmann 2010a	0.87	**1.57**	**1.82**	1.94	2.24	0.89	1.76	2.17
MYCN	2p24.3	McElwaine 2004	0.35	1.00	2.87	0.77	2.19	**0.34**	2.99	2.28
PRAME	22q11.22	McElwaine 2004	1.60	11.59	7.24	15.61	9.76	**4.18**	2.77	3.73
SMOX	20p13	Ge 2006	**1.37**	0.76	0.55	1.02	0.75	0.79	0.96	1.30
ST18	8q11.23	Klusmann 2010b	0.90	0.74	0.82	0.74	0.82	0.78	0.96	0.95

Expression values of genes known for their role in MK differentiation and DS or non-DS AMKL development in each of the eight comparisons: DS AMKL (pool 'A') vs. non-DS AMKL (pool 'B'); DS AMKL (pool 'A') vs. normal MK (pool 'D'); non-DS AMKL (pool 'B') vs. normal MK (pool 'D'); DS AMKL (pool 'A') vs. normal CB MK (pool 'E') cells; non-DS AMKL (pool 'B') vs. normal CB MK (pool 'E'); DS AMKL (pool 'A') vs. TMD (pool 'C'); TMD (pool 'C') vs. normal MK (pool 'D'); TMD (pool 'C') vs. normal CB MK (pool 'E'). Data extracted from the full tables showing expression values and their ratio for about 17-26,000 loci for each comparison. In bold: expression ratio values are consistent with data available in the literature (see references indicated). The other values are reported for completeness. Descriptions of the gene symbols are given in Additional file 2 (available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/).

4.3 HR-DSCR on human chromosome 21 and gene hunting

4.3.1 HR-DSCR computational biology analysis

BLASTX research did not find a significant similarity with the amino acid sequences known in the region between 37,929,000-37,975,000 bp. Comparing the sequence of interest with the EST sequences cataloged in the GenBank database, there was no evidence of the presence of EST sequences matching the HR-DSCR.

The FGENESH program shows the presence of a putative gene on the complementary strand consisting of 7 exons. The predicted amino acid sequence does not have homology with the amino acid sequences of known proteins, neither human nor of other organisms. Having found no prediction about the presence of genes encoding for protein on the reference filament, we focused on the research of transcripts with regulatory function. The miRBase (Kozomara and

Griffiths-Jones, 2014) analysis of the region NC_000021.9 between 37,959,229-37,963,130 confirmed a significant homology between a stretch of about 100 bp and the known miRNA: hsa-mir-1273a (Figure 11).

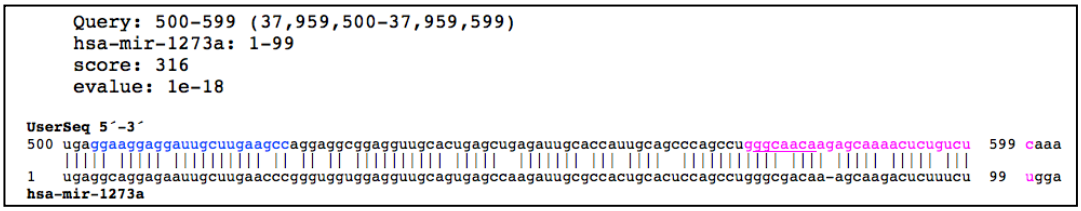


Figure 11 miRBase analysis of the 1,000 bp sequence on the NC_000021.9 between 37,959,001-37,960,000 bp.

4.3.2 Putative miRNA1 molecular analysis

4.3.2.1 Northern Blot (miRNA1 probe)

The labelling of the miRNA1 probe, used for hybridization on the MTN2 filter, was measured at the beta-counter, which took over a radioactivity of 1,093,898 counts per minute (cpm) in 1 µL. After the hybridization step, we proceeded with the exposure of the filter autoradiographically and subsequently developed in the darkroom. The development of the autoradiographic sheet made clearly visible two bands at the same height but in two different tissues, such as heart and skeletal muscle (Figure 12).

We measured the distance between the visible bands on the sheet and the loading wells and we reported the measurements in logarithmic scale on graph paper. Taking advantage of the straight line for the MTN2 filter, according to the method described in the Materials and Methods section, it was possible to convert the distance in millimeters in the number of nucleotides. Both bands correspond to a transcript of approximately 13.5 Kb (Figure 12).

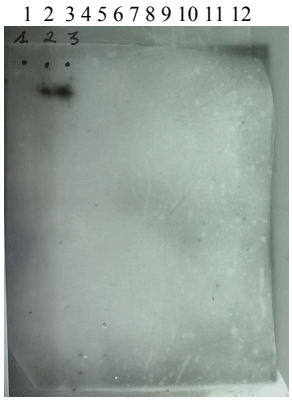


Figure 12 Autoradiographic sheet developed after 100 h of exposure to the commercial filter MTN2 hybridized with the probe for the putative miRNA1. Lane 2: total RNA of heart; lane 3: total RNA of skeletal muscle. In the other lanes the following total RNA are present: 1) brain, 4) colon, 5) thymus, 6) spleen, 7) kidney, 8) liver, 9) intestine, 10) placenta, 11) lung, 12) peripheral blood leukocytes.

Exposure to the phosphor screen was acquired at PhosphorImager. The result was the same as the autoradiographic sheet (Figure 13).

Figure 13 Capturing of PhosphorImager of phosphor screen after 24 hours of exposure to the commercial filter MTN2 hybridized with the probe for putative miRNA1.

Exposure to the phosphor screen was also performed after washing the filter necessary to remove the probe remained linked. The acquisition at the PhosphorImager showed that the washing is successful since the hybridization signal of the probe was no longer visible.

4.3.2.2 Northern Blot (pri-miRNA1 probe)

Following the result described above we have designed a probe specific for a region longer than the previous examined. This probe is complementary to the sequence of the primary form of the putative miRNA1 and within the limits 37,959,623-37,959,673. The labelling of the pri-miRNA1 probe, used for hybridization on the MTN2 filter, was measured at the beta-counter, which took over a radioactivity of 525,219 counts per minute (cpm) in 1 μL. After the hybridization step, we proceeded with the exposure of the filter to the autoradiographic sheet subsequently developed in the darkroom. The development of the autoradiographic sheet made clearly visible two bands at the same height (13.5 Kb) of those obtained in the first hybridization with the miRNA1 probe, in the same tissues such as heart and skeletal muscle (Figure 14).

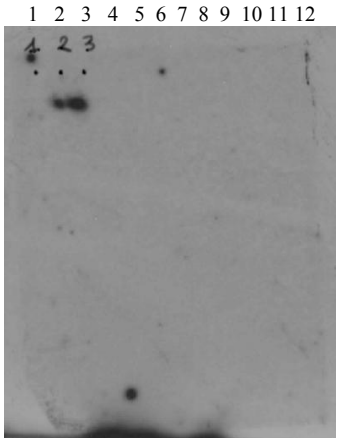


Figure 14 Autoradiographic sheet developed after 72 h of exposure to the commercial MTN2 filter hybridized with the pri-miRNA1 probe. The hybridization signals are in lanes 2 and 3 and correspond, respectively, to total RNA of heart and skeletal muscle.

4.3.3 *DSCR4* locus genomic organization

The analysis on UCSC Genome Browser of the *DSCR4* locus found the presence of a EST sequence aligned in part to the HR-DSCR with the identification code CB993317.

We performed an analysis using the software NCBI Splign (http://www.ncbi.nlm.nih.gov/sutils/splign/) of this EST to identify possible splicing site against the entire genomic sequence of human chromosome 21. We obtained the alignment of EST to 4 exons which identify a new isoform of the gene *DSCR4* not described. This isoform, referred to conventionally as *DSCR4.2*, has exons 1 (E1) and 2 (E2) in common with the isoform 1, while differs for the lack of exon 3 (E3) and having the last two exons in 3' end, which are then classified as exon 4 (E4) and exon 5 (E5) (Figure 15).

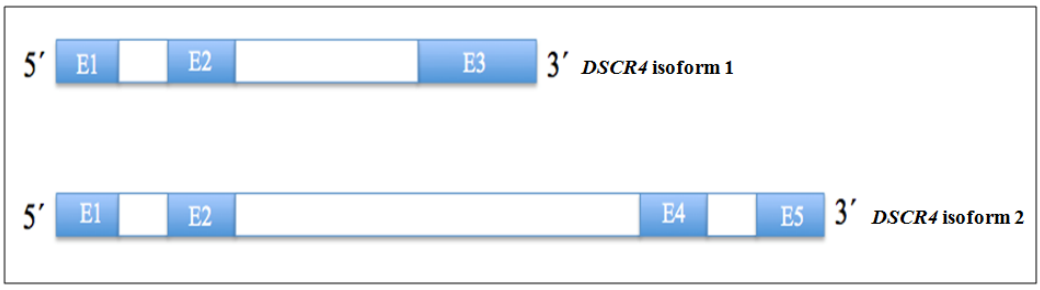


Figure 15 Genomic organization of locus *DSCR4*. Explanation in the text.

The data recorded in the NCBI database of the EST CB993317 show that it was identified from a cDNA library made from placental tissue. To confirm the genomic organization of the identified transcript, the relative electropherogram was obtained through the database UniGene (http://www.ncbi.nlm.nih.gov/unigene).

The study of this circuit detects an excellent quality of sequencing, allowing the unequivocal confirmation of the two splice junctions E2-E4 and E4-E5 (Figure 16).

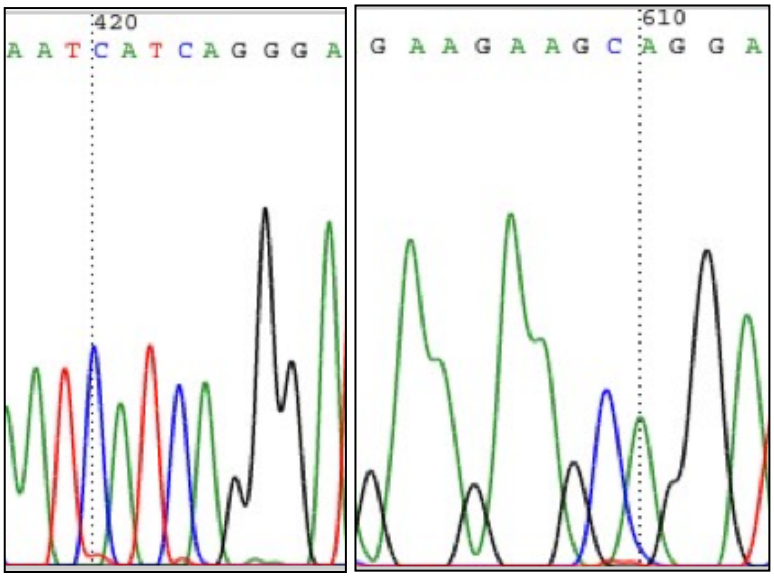


Figure 16 Electropherogram of the EST sequence CB993317 illustrating the 6 bases just before and after the splicing sites E2-E4 (left) and E4-E5 (right).

The exact boundaries of the exons (E) and introns (I) of the specific isoform 2 are given in Table 36. It is inferred that the overall extension of the *DSCR4* locus is of 169,936 bp, due in particular to the addition of a very long intron. Compared to the previous estimate of 67,350 bp, the new model of the locus *DSCR4* is increased by 102,586 bp.

	Beginning	**End**	**Dimensions bp**	**Beginning CDS**	**End CDS**	**CDS bp**
DSCR4.2 E1	38,121,360	38,121,128	232	38,121,255	38,121,128	127
DSCR4.2 I1	38,121,127	38,120,409	718			
DSCR4.2 E2	38,120,408	38,120,307	101	38,120,408	38,120,307	101
DSCR4.2 I2	38,120,306	37,953,006	167,000			
DSCR4.2 E4	37,953,005	37,952,821	184	37,953,005	37,952,852	153
DSCR4.2 I3	37,952,820	37,951,702	1,118			
DSCR4.2 E5	37,951,701	37,951,425	276			

Table 36. Exons and introns coordinates of the isoform *DSCR4.2* (based on GenBank sequence NC_000021.9). bp= base pairs; CDS= coding DNA sequence.

It is therefore legitimate to reconstruct the overall sequence of the new transcript identified, leading to the prediction of a sequence coding for a protein, in particular for a polypeptide of 127 amino acids (Figure 17).

```
  1 MSLIILTRDDEPRIFTPDSDAASPALHSTSPLPDPASASPLHREEKILPKVCNIVSCLSF
 61 SLPASPTDSGLASPTIIRDPVASKSQLELSSRPAKTEPRAWSAQFGRCSAQPQPQVDFLP
121 AENTCSG
```

Figure 17 Sequence of the predicted amino acid product of *DSCR4.2* transcript. The specific amino acids of this isoform are shown in blue.

The systematic search of homology with other proteins using PSI-BLAST is not allowed to highlight any significant similarity to any polypeptide or known protein family, nor to functional domains. The polypeptide is identical to *DSCR4* for the first 77 amino acids, then by departing from *DSCR4* in the carboxyl-terminal region of the sequence.

4.3.4 *HRDSCR1* Locus

The results described above allow to identify and provide an initial description of a new complex locus contained in the HR-DSCR, which is conventionally called here *HRDSCR1*. The structure of this locus is derived from the demonstration that the stretch of about 100 bases identified as putative miRNA is actually transcribed and included in a long transcript of approximately 13 kb expressed in skeletal muscle and human heart. In addition, the bases present on the complementary strand are expressed as part of a large intron, which we identified as belonging to a new isoform of RNA transcribed from the known locus *DSCR4* (Figure 18).

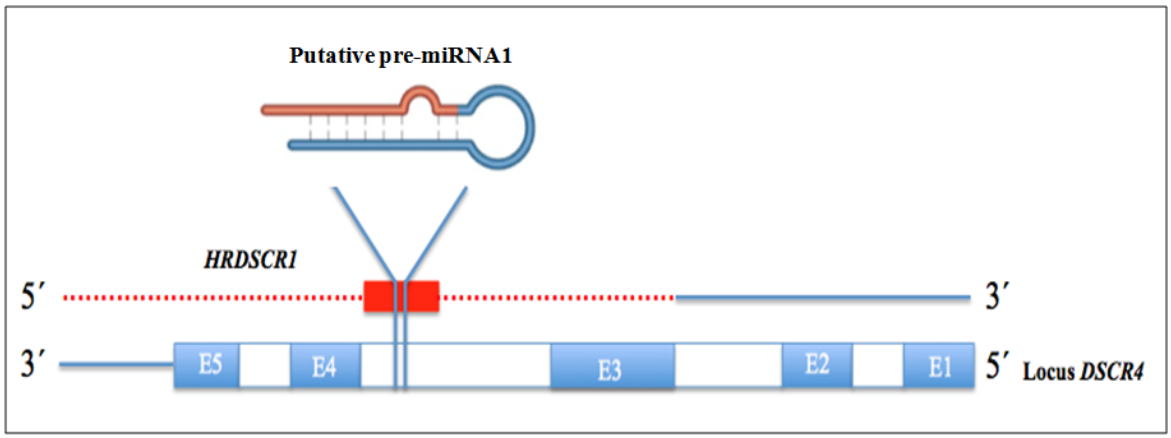


Figure 18 Representation of the region of human chromosome 21 containing the two loci *HRDSCR1* and *DSCR4* which have an opposite orientation. On strand 5'-3' we find the locus *HRDSCR1* and on the complementary strand the locus *DSCR4*. Highlighted the region common to the two loci containing on the strand 5'-3' the putative miRNA1 precursor form.

From these data we can deduce that the 37.9 Mb region included in the HR-DSCR is not intergenic, as one might assume from current maps of human chromosome 21, but is transcribed and constitutes a new human locus. This locus appears bidirectionally transcribed leading to the formation of two partially overlapping transcripts at the level of the region of the putative miRNA1.

Chapter 5

Discussion and conclusions

5.1 Systematic meta-analysis of gene expression profiles of normal human tissues

In this work we proposed an integrated model of the human transcriptome of the whole brain (adult and foetal) and three of the brain regions severely affected in Down syndrome (DS), such as the cerebellum, cerebral cortex and hippocampus (Haydar and Reeves, 2012), and of the whole heart.

TRAM software allowed us to integrate data from different sources, including data from different microarray experiments, also performed on different platforms through a method of intra-sample and inter-sample normalization (quantile scaled normalization). We used a set with a relevant number of samples to create a quantitative transcriptome reference map for each of tissues and organs. This map allowed us to identify critical genome regions which are over/under-expressed in normal brain and heart and are typical of these tissues compared to the pool of all the other human tissues. It also allowed us to assign a linear reference expression value to each locus.

5.1.1 Whole normal human brain transcriptome map

The adult brain transcriptome map showed the highest expression value of a segment on chromosome 12, including the known genes *TUBA1A*, *TUBA1B* and *TUBA1C* (Baumann et al., 1996; Okumura et al., 2013). This finding appears to be an evident genetic correlate of the well known main structural characteristic of the nervous system functional unit, the neuron. The cytoskeleton is most responsible for the neuron shape and function, it consists of microtubules and the expression of tubulins is typically used as a marker of neural differentiation. The alpha and beta tubulins represent the major components of microtubules. The analysis at single gene level showed that the first over-expressed gene is *BCYRN1*, while *TUBA1B* is the second. *BCYRN1* encodes a small neural non-messenger RNA, the sequence contains a 5′ portion homologous to the Alu Lm (left-hand Alu monomer) and seems to have a regulative function (Wang et al., 2002; Mus et al., 2007). This RNA probably belongs to the subfamily of Alu-elements over-represented in the transcriptome of brain tissue (Faulkner et al., 2009; Xu et al., 2010). As pointed out by Lejeune (1988), some studies supported the hypothesis that deficiency in the neurotubule network could be associated with mental disorder (Nunez, 1985). The relationship between some mental disorders and increased levels of *BCYRN1* was demonstrated. Its over-dosage seems to be associated with the increase of the cytoskeleton fibres causing a blockage in transport of the RNA within the cell due to the fibre overcrowding (Mus et al., 2007). Among chr21 genes, *SOD1*, *APP* and *DNAJC28* have the highest expression values. A

preclinical research study conducted on mice used both *SOD1* and *APP* as therapeutic targets in support of the hypothesis of the over-dosage effect of chr21 genes in the pathogenesis of Down syndrome (Costa and Scott-McKean, 2013). The *DNAJC28* gene has not yet been characterized and its function could be interesting to investigate.

To highlight the brain-specific gene expression profile, a comparison pool composed of 622 datasets derived from 53 different human tissues or organs has been accurately assembled. Spinal cord samples (n=24) have been included in this pool, considering the distinctive anatomical and functional features of the brain and spinal cord, and the extensive overlapping of results when the spinal cord samples are included in or removed from the non-brain sample pool.

When we compare the adult brain gene expression with the pool of tissues minus brain gene expression, the highest statistically significant expression value is attributed to a segment on chromosome 5, including the known genes *GABRA6*, *GABRA1* and *GABRG2* encoding for subunits of the GABAergic receptor, one of the major inhibitory receptors in mammalian brain. An altered expression of GABA receptors is typical in the brains of subjects with autism and Fragile X syndrome, another common form of inherited mental retardation (D'Hulst et al., 2006; Fatemi et al., 2009). The significantly under-expressed segment is on chromosome 2 and contains *REG1B*, *REG1A* and *REG3A* genes, encoding for proteins secreted by the pancreas (Sanchez et al., 2001; Zhou et al., 2010; Christa et al., 1996). These data allow the affordable identification of genes expressed specifically in the brain and not in other tissues.

At single gene level, *ANKRD30B* is over-expressed in the adult brain compared to the pool of non-brain tissues. It is also known as the *NY-BR-1.1* gene, a homologous gene of *NY-BR-1*, with which it shares 54 % of amino acids, but unlike *NY-BR-1* gene expressed in breast and testis, it is also expressed in the brain. The ankyrin proteins carry out a wide variety of biological activities, they have a repeat motif recognized in more than 400 proteins, including cyclin-dependent kinase inhibitors, transcriptional regulators, cytoskeletal organizers, developmental regulators, and toxins (Jäger et al., 2001).

Among chr21 genes, *DNAJC28*, *LINC00320*, *LINC00323*, *OLIG1* and *OLIG2* are over-expressed. *OLIG1* and *OLIG2* are known to have the function of transcription factors in oligodendrocytes, cells that perform the main function of myelination of neurons in the central nervous system. A recent study suggested a link between the over-dosage of *OLIG2* in trisomy 21 and the reduced brain size in affected individuals (Lu et al., 2012). Instead, *LINC00320* and *LINC00323* are non-coding sequences that may have a regulatory role and that seem to have a fundamental role in the adult brain because they maintain a low expression value in the foetal brain transcriptome map. Their function could be interesting to investigate.

The foetal brain transcriptome map, as the adult brain transcriptome map, showed the highest expression value of the segment on chromosome 12, including the known genes *TUBA1A*, *TUBA1B* and *TUBA1C*. It is interesting to note that at two very different developmental stages of the human brain, the same chromosomal segment (12q13.12) is over-expressed. At single gene level, *TUBA1B* has the highest expression value. Among the genes of chr21, *DNAJC28*, followed by *SOD1* (1,657.36) and *ATP5O* are over-expressed. In this case we note the same *DNAJC28* and *SOD1* expression patterns between adult and foetal brain, while it is interesting to note that *ATP5O* is over-expressed.

Comparison between the foetal brain and the adult brain transcriptome map showed the highest expression value of a segment on chromosome 8, including the known gene *BHLHE22* and the uncharacterized loci, *LOC100130155* and Hs.388788. *BHLHE22* gene encodes for a transcription factor that regulates cell fate determination, proliferation, and differentiation (Xu et al., 2002). It seems to be involved in neuronal development rather than in the functions performed by the adult brain. *LOC100130155* and Hs.388788 are loci which have not yet been characterized but are probably fundamental in brain development.

At single gene level there is a high expression ratio for *TMSB15A* gene and for *DCX* gene, which have an increase of 63.15 times and of 42.79 times, respectively. The former is involved in cellular motility (Bao et al., 1996; Gu et al., 2008). Its expression profile was analyzed in normal and pathological tissues (Yokoyama et al., 1996) and it was considered a neuroblastoma specific gene. The derivation of neuroblastome from neuroblasts or nerve undifferentiated cells allows us to draw the further conclusion that neuroblastoma specific genes might include genes that are especially expressed in the developing nervous system. The *DCX* gene is involved in neuronal migration during development by regulating the organization and stability of microtubules (Fung et al., 2011; Bechstedt and Brouhard, 2012). It was demonstrated that *DCX* mutations are associated with disorders caused by an abnormal migration of neurons during development (Hehr et al., 2007). The low expression ratio of *LINC00320* could be significant of the regulatory role played by this locus in the adult brain only. This information favours the hypothesis that the foetal brain transcriptome map can be used as a reference for the rapid identification of genes involved in the development of the central nervous system.

We noted that many differences exist between the gene expression patterns of the adult and foetal brain, indeed many adult brain genes are not yet active at such an early phase of development as the foetal stage, while many foetal brain genes which are less active in the adult brain are possibly those involved in brain development.

The experimental screening was performed on the adult brain transcriptome map through a "Real-Time" RT-PCR experiment using 9 genes selected from the brain TRAM map (see

"Materials and Methods"). One of these genes is chosen as reference, according to its medium-high expression value among the 9 genes selected. The observed and expected expression ratios are listed in Table 16. It is interesting to note that the correlation between the values calculated by TRAM using 60 normal human brains and an experimental comparison using independent samples is maintained through several orders of magnitude and from the lowest to the highest values.

The significant correlation indicates that the alternative method of sample collection – a different tissue source of the sample and the different post-mortem interval of tissue retrievement – was neutralized by the high number of samples. In fact, as regards the post-mortem interval (PMI), it was proven that undegraded mRNA may be isolated from most brain regions many hours postmortem (up to 30 hours), so the PMI is not actually predictive of mRNA integrity (Ervin et al., 2007).

Although there are studies showing a sex-biased brain gene expression (e.g., Trabzuni et al., 2013), it was also proven that gender does not have a significant influence on the observed gene expression changes, at least when a significant number of samples is available for study (Lockstone et al., 2007). Following the observation by Trabzuni and coll. (2013) that a minority of genes (2.6%) presents sex-biased expression level in specific brain subregions, it is expected that differential gene expression analysis between male and female whole brains affects a small number of genes. Table 5 shows the first ten over-/under-expressed genes identified by our analysis. For several of them we have found literature evidence about their relationship with gender (e.g., *DRD4* (Vawter et al., 2004; Dmitrieva et al., 2011), *FTHL17* (Wang et al., 2001), *XIST* (Brown et al., 1992). *DRD4* gene, encoding for D4 subtype of the dopamine receptor, seems to be involved in the difference in the number of dompaminergic neuron between males and females (Vawter et al., 2004). *DRD4* gene has been extensively studied for its possible association with increased vulnerability to antisocial and impulsive behaviour, including behaviour recurrent in males and not in females (e.g., Dmitrieva et al., 2011). *FTHL17* gene, encoding for ferritin, heavy polypeptide-like 17, was described as a typical example of X-linked genes expressed only in male germ cells (Wang et al., 2001). It could be interesting to investigate its linkage with brain tissue. *XIST* gene, encoding for X inactive specific transcript (non-protein coding) (Brown et al., 1992) has the lowest expression ratio between male and female brain in the differential transcriptome map, demonstrating that it is a gene typically expressed in females. These results are a confirmation of the reliability of the data produced by our map.

For the others genes shown in Table 5, it would be interesting to investigate their relationship with gender, in particular for the neuronal gene *NPAS4*, encoding for neuronal PAS

domain protein 4 and seeming to be involved in the transcriptional regulation of some neural genes in the brain (Ooe et al., 2004), and for *CCL3*, *IL1B* and *CCL4* genes encoding respectively for chemokine (C-C motif) ligand 3, interleukin 1, beta and chemokine (C-C motif) ligand 4, all three of which are involved in immune system responce (Guo et al., 2003; Zunszain et al., 2012).

In summary, the high correlation (r=0.94, data not shown) between the male- and female-derived gene expression values support the utility of a human brain map as a general standard reference, while the creation of differential expression maps comparing male- and female-derived samples remains feasible using our approach, and may highlight sex-biased differences for specific genes, whether they are located on the sex chromosomes or not.

The experimental validation of our map allows us to consider all the other intervening values among the points we selected as *bona fide* values, thus for the first time providing a complete and quantitative representation of the reference expression values for 39,250 mapped and 26,026 unmapped transcripts of the normal human brain.

The initial analysis of the biological meaning of these data clearly shows a relationship of our findings with what is already known about genes which have either a relevant function or are almost silenced in the brain. While the correlation between the values obtained by meta-analysis of multiple gene profiling datasets and independent samples assayed by a different method (Real-Time RT-PCR) is excellent, some remaining differences may easily be attributed to the physiological biological differences among the samples, as well as to experimental error. In some cases, specific differences in the method of analysis can explain the variability of the expression value. For instance, *BCYRN1* contains a 5′ portion homologous to the Alu Lm, and the corresponding array probe (37 bp) recognizes this domain. So it is possible that, in addition to *BCYRN1*, other members of this family of interspersed repetitive DNA (Martignetti and Brosius, 1993) hybridized to the probe. This could explain a hybridization signal higher than that observed, indeed the observed gene expression relative to the reference gene is 7.11 by RT-PCR and 12.8 by meta-analysis of cDNA hybridization microarrays.

5.1.2 Cerebellum transcriptome map

The cerebellum transcriptome map showed the highest expression value of a segment on chromosome 11 containing an EST cluster and the long non-coding RNA (ncRNA), *MALAT1* (Kryger et al., 2012). This ncRNA is very abundant in neurons where it controls the expression of a subset of genes significantly involved in nuclear and synapse function and regulates synaptogenesis (Bernard et al., 2010). At single gene level *BCYRN1* is over-expressed as in the

whole brain, while among chr21 genes, *DNAJC28* and *SOD1* remain the genes with the highest statistically significant expression values.

When we compare the cerebellum gene expression map with the adult brain gene expression map, the highest value is attributed to *CBLN3*. It is a new member of the precerebellin family coding for a neuropeptide detected in the granule neurons of cerebellum that, together with CBLN1, regulate synapse integrity and plasticity in Purkinje cells (Iijima et al., 2007). The lowest expression value is attributed to *NRGN*. This gene has specific functions in the cerebral cortex (Li et al., 2003), it is a target of thyroid hormone and in hypothyroidism conditions suffers an altered control of expression which causes mental disorder during development (Shen et al., 2012). The experimental validation of the expected data was made on 8 genes of the whole cerebellum transcriptome map, again finding a higly significant statistical correlation. We noted that the observed expression ratios of *RCAN1* and *GPCPD1* indicate their over-expression in comparison with the reference gene instead of being under-expressed as the data anticipated. This result could be due to a SD as a percentage of expression calculated by TRAM >95 for *RCAN1* and *GPCPD1*. A high SD could explain the significant distance between expected and *in vitro* observed results, indeed the exclusion of *RCAN1* and *GPCPD1* from the statistical analysis determines the increase of the statistical correlation from 0.96 to 0.97.

5.1.3 Cerebral cortex transcriptome map

In the cerebral cortex TRAM map the significantly over-expressed segment is again on chromosome 12, including the over-expressed known genes *TUBA1A*, *TUBA1B* and *TUBA1C*. At single gene level, the EST cluster Hs.732685 is the first over-expressed locus, followed by *TUBA1B*. Among chr21 genes *OLIG1*, *PCP4* and *SOD1* have the highest statistically significant expression values. *PCP4* is known to be highly expressed in the brain, primarily in the Purkinje cells, a class of GABAergic neurons located in the cerebellar cortex (OMIM entry #601629).

When we compare the cerebral cortex gene expression with the adult brain gene expression, the over-expression of the *KRTAP13-2* gene encoding for a protein associated to keratin, is interesting. In contrast, *DNAJC28* and *LINC00320* are under-expressed: the only time this happens to *DNAJC28* in our maps. The experimental validation of the expected data was made on 8 genes of the whole cerebral cortex transcriptome map. The gene expression value as calculated by TRAM for *GAPDH*, *NCDN* and *RCAN1* was not observed. The first two have a SD as a percentage of expression calculated by TRAM >95, while the third has a SD of 90 %. The statistical correlation between expected and observed ratios is not significant. To improve the correlation, we excluded those genes with a SD as percentage of expression value >95 from the

statistical analysis and, in fact, after this selection, the correlation became highly significant.

Applying the same criterion to the previous maps which were experimentally screened, the correlation remained the same in the whole brain transcriptome map (Figure 3b), whereas it further increased in the whole cerebellum transcriptome map (Figure 4b). That the significance was maintained confirms the theory that our maps could be used as a reference for expression studies performed on whole brain.

5.1.4 Hippocampus transcriptome map

The hippocampus transcriptome map showed the highest expression value of a segment on chromosome 12 (12q13.12) (Table 6), including the over-expressed known genes *TUBA1A*, *TUBA1B* and *TUBA1C*. The over-expression of this segment was already found in the whole brain transcriptome map (Caracausi et al., 2014). It describes the main structural feature of the nervous system functional unit, the neuron, characterized by the preponderant extension of the neurotubules of its cytoskeleton. This segment returns as the fourth over-expressed, including two other genes: *ARF3* and *DDN*. It is worth the presence, together with the tubulin genes, of the *DDN* gene encoding for a protein which interacts with the cytoskeleton components and seems to modulate the structure of the postsynaptic cytoskeleton (Kremerskothen et al., 2006). The second over-expressed segment is on chr8 (8p21.2) (Table 6), including the known genes *BNIP3L*, *PNMA2* and *DPYSL2*. This genome region was found to have links with the severe mental disorder schizophrenia (SZ) (Fallin et al., 2011). This is a gene-rich region which includes 32 annotated genes, among them the strongest statistical evidence of association with SZ is for *DPYSL2* (Fallin et al., 2011). In mammals, *DPYSL2* is required to induce axonal outgrowth, maintaining neuronal polarity and promoting microtubule assembly in hippocampal neurons. It has been implicated in multiple neurological disorders associated with impaired neuronal plasticity and neurodegeneration (Czech et al., 2004; Williamson et al., 2011).

A segment representation of the genome could help to identify chromatin domains whose genes are encoding for proteins interacting with each other or belonging to the same cell pathway. This is the case of the over-expressed segment on chrX, including the genes of the BEX (brain-expressed X-linked) family: *BEX1*, *BEX2* and *NGFRAP1* (Table 6), which are known to be involved in the same neuronal signaling pathway (Naderi et al., 2007).

At single gene level, the over-expression of an EST cluster, Hs.732685 (mapping within the known locus *MALAT1*, as showed in the Results section), and of the known gene *CALM2* followed by *TUBA1C* and *TUBA1B* emerges (Table 7). CALM2 is an intracellular Ca2+-binding protein found mainly in the central nervous system (Shirasaki et al., 2006). It mediates many

physiological functions in response to changes in the intracellular Ca2+ concentration, such as neuronal cell death due to Ca2+ overload in neurons after excitotoxic insult. It was reported that calmodulin antagonists protect hippocampal CA1 neurons against hypoxia and attenuate brain damage after transient focal ischemia (Sun et al., 1997; Sato et al., 2003). *TUBA1A*, *TUBA1B* and *TUBA1C* are among the first ten over-expressed genes of the hippocampus transcriptome map. Alpha and beta tubulins are typically used as markers of neural differentiation (Kumar et al., 2010; Okumura et al., 2013). They are needed for neurotubule network assembling, which is fundamental for neuron shape and function and seems to be disturbed in some mental disorders (Lejeune, 1988; Baumann et al., 1996).

Due to the fact that the hippocampus is one of the main brain sub-regions which is severely affected in DS (Bartesaghi et al., 2011), particular attention was paid to the expression of chr21 genes in this brain region. Among the chr21 genes, the over-expression of *SOD1* and *OLIG1* emerges (Table 7). They typically have a high expression pattern in the whole brain (Caracausi et al., 2014). Ahead of them, the long intergenic non-protein coding RNA, *LINC00114*, also known as *C21orf24*, is over-expressed. It was previously identified following a study of detailed annotation of a genomic DNA region (21q22.2-q22.3) on human chromosome 21 (Owczarek et al., 2004). The characterization of *C21orf24* identified six different mRNA alternative splice forms, and the failure to identify any significant open reading frames in all the alternative splice forms suggests that it should be classified as a gene for non-coding RNAs (Owczarek et al., 2004). The *LINC00114* sequence turns out to be identical to the *C21orf24* isoform 3, as shown by a BLASTN analysis. The relational database at the core of the TRAM software allowed us to track the probes for *LINC00114* used by the microarray platforms included in our analysis, and showed that they actually are identical to *C21orf24* isoform 6. We tried to clone this sequence and found that it is not over-expressed as it appears to be from the microarray hybridization signal (data unshown). We did not find repeated sequences in the region recognized by the probes, so this finding remains unexplained.

The differential transcriptome map obtained by the comparison of the hippocampus vs. the whole brain transcriptome map again showed, at single gene level, the over-expression of the chr21 locus, *LINC00114*, suggesting a typical hippocampal function. Among the other over-expressed genes there are *SLC44A4* and *IL10* genes (Table 8). SLC44A4 belongs to a new family of choline transporters, the choline transporter-like proteins (CTL1-5). A recent study showed its functional linkage to acetylcholine synthesis and secretion in non-neuronal cells (Song et al., 2013), even if our data show its exact over-expression in the brain sub-region. Interleukin-10 (IL-10) is a cytokine with anti-inflammatory properties which negatively modulates proinflammatory cascades at multiple levels in the central nervous system. Cytokines

are involved in bidirectional signaling between the central nervous system and the peripheral immune system, and play a role in cognitive processes. Cytokines also take part in the regulation of neurogenesis: the proliferation of new neurons which is crucial for hippocampal functions such as learning and memory (Arisi, 2014).

Emerging as one of the most expressed chr21 genes is the *AIRE* gene (Table 8), encoding for an autoimmune regulator performing a fundamental function in the thymus. Its expression was studied in DS individuals whose immune phenotype is characterized by thymus hypotrophy (Lima et al., 2011). It is under-expressed in DS individuals compared to normal individuals, probably due to a down-regulation of this gene among the chr21 genes as an effect of the genomic-dosage imbalance (Prandini et al., 2007).

Analysis of the differential transcriptome map also allowed us to identify the genes with a low expression level in the hippocampus, which probably do not have a typical function in this brain structure when compared to the whole organ. For example, the under-expression of *DNAJC28* gene (Table 8), previously found to be a typical brain gene compared to a pool of non-brain tissues (Caracausi et al., 2014), may imply that the hippocampus is not the cerebral sub-region in which it performs its function. In addition, this gene has a high expression level in the foetal brain and in the cerebellum vs. whole brain but, in the cerebral cortex only, it has a 50-fold decrease (Caracausi et al., 2014). These data confirm that the brain is a very complex organ and that a reference transcriptome map of its sub-regions is necessary in order to piece together how they individually function.

Studies showing a sex-biased brain gene expression (Trabzuni et al., 2013) led us to investigate whether gender has a significant influence on the hippocampus gene expression profile observed in our analysis. We performed an additional differential transcriptome map to investigate specifically sex-biased gene expression patterns.

The relationship between gene expression profile and gender was found for the genes included in the genomic segments which were significantly over-/under-expressed in the differential transcriptome map. The first over-expressed segment is on chrY (Table 9) and includes *TTTY15*, *USP9Y* and *DDX3Y*, all encoding for products involved in spermatogenesis and having at single gene level an expression ratio higher than 2. The first under-expressed segment is on chromosome X and includes the known genes *XIST*, *JPX* and *FTX*. All deriving products are involved in the chrX inactivation process and their expression ratio is lower than one. *XIST* in particular has a 2.8-fold decrease, confirming that it is a typically expressed gene in females. These results are an additional confirmation of the reliability of the data produced by our map.

At single gene level, the highest expression ratios were observed for genes that have a typical function in the muscles (Table 10). The sex-biased expression of these genes suggests that they could behave like tissue-indipendent gender-specific transcripts (Staedtler et al., 2013).

To be sure that gender did not lead to a significant global gene expression variation, we performed a statistical bivariate analysis between male and female gene expression values. This showed a significant correlation of data ($r=0.98$, p-value<0.0001) (Figure 2), probably because, when a great number of samples ($n=41$) is available, the numerous demographic differences do not affect this kind of analysis (Lockstone et al., 2007). In agreement with a recent study by Trabzuni et al. (2013) showing that a small number of genes (2.6%) presents sex-biased expression levels in specific brain sub-regions, we noted that *XIST* – specifically activated in female cells to start the X-inactivation process – is the most differentially expressed chrX gene in female (value=606.58) and male (value=215.35) hippocampus (ratio=2.8). In contrast, genes located in the pseudoautosomal region of chrX – which are known to escape X-dosage compensation and to be represented in two active copies in both female and male cells – appear to be expressed at very similar levels (for example, ratio=1.23 for *ASMT* and ratio=1.16 for *ARSD*).

To prove the reliability of the hippocampus transcriptome map generated by TRAM software, we performed an experimental validation of the obtained data. The significant correlation between the expected and observed data ($r=0.99$, p-value<0.0001) (Figure 6) allows us to consider all the other intervening values among the points we selected as *bona fide* values, thus for the first time providing a reliable and quantitative representation of the reference expression values for 30,739 mapped transcripts of the normal human hippocampus.

Initial analysis of the biological meaning of these data clearly shows a relationship of our findings with what is already known about genes which have either a relevant function or are almost silenced in the hippocampus.

While the correlation between the values obtained by meta-analysis of multiple gene profiling datasets and independent samples assayed by a different method (Real-Time RT-PCR) is excellent, some differences between the two types of data are not to be excluded, due to the physiological and biological differences among the samples, experimental errors during microarray analysis, or differences in the method of analysis.

The availability of the systematic and detailed expression map presented here for the hippocampus represented an excellent occasion to investigate, among many features of the transcriptome, the suitability of individual genes as the best stable reference genes for data normalization in gene expression studies. This is because they fulfill criteria including a widely diffused, constant and high expression (Casadei et al., 2011). The results offer a selection of

many known genes and uncharacterized loci, mostly EST clusters (Supporting Information Table S5, available at: http://apollo11.isto.unibo.it/suppl/2015_Caracausi/). We would like to enfasize that *TBCB* is the best gene at behaving like a housekeeping gene (Supporting Information Table S5, available at: http://apollo11.isto.unibo.it/suppl/2015_Caracausi/). It encodes for a tubulin cofactor which is fundamental for the polymerization of microtubules, the main components of which are over-expressed in our transcriptome map. It is also involved in microglial cytoskeletal changes during microglia transition in response to injury or pathogen invasion of the central nervous system (Fanarraga et al., 2009).

Until now, no study had been carried out by relating such a large number of samples, integrating data from several experiments conducted on different platforms, and providing information about 39,250 loci in the adult brain map, about 38,163 loci in the cerebellum map, about 27,504 loci in the cerebral cortex map and about 30,739 loci in the hippocampus map. The new TRAM version issued includes updated gene tables of Entrez Gene and UniGene, improving gene localization data and parsing. It solves the problem of many aliases of each gene, allowing the correct assignment of each probe to the loci for known transcripts and EST clusters. The EST clusters and the uncharacterized loci mentioned in our work would be interesting to investigate. Their localization was derived from UCSC "ESTs" track in the UCSC Genome Browser (Kuhn et al., 2009), which was also imported and processed during the TRAM set-up (Lenzi et al., 2011).

There are several databases and atlases of brain gene expression with specific features publicly available online, for example GeneCards (Safran et al., 2010), Allen Atlas (Hawrylycz et al., 2012) and BrainSpan Atlases (http://www.brainspan.org/). GeneCards is an integrated database that provides genomic, transcriptomic, proteomic, genetic, clinical, and functional information on human genes. However, it provides transcriptomic information that comes from a limited number of hybridization experiments performed only on the Affymetrix GeneChip HG-U95 set. Furthermore, this database does not give information about uncharacterized loci, such as EST clusters. Allen Brain Atlas is a transcriptional atlas of the adult human brain comprising microarray profiling performed on macrodissected and on laser microdissected brain regions. It provides expression data on brain specific regions. It provides a very useful amount of data for the experimenter but remains incomplete when searching for data derived from whole tissue and from a large number of samples. Furthermore, the number of loci and the number of hybridizations per single locus are lower than those analyzed by TRAM software. This is confirmed by the data attributed to *SOD1* expression in the brain. Although it is possible to see the expression of *SOD1* for each macrodissected brain region, it is not possible to have a reference expression value for the entire tissue. The hybridization signal is only provided by two

probes which are both from the same platform, whereas the one provided by our transcriptome map was obtained from the normalization of 96 data point expression values and 9 different platforms. The TRAM data for *BCYRN1*, the gene with the highest expression value in the brain, is not searchable in the Allen Atlas, probably due to the low number of genes about which it gives information, and the same goes for clusters of EST. The different type of data does not allow us to make a comparison between the expression values provided by TRAM and those provided by the Allen Atlas. BrainSpan Atlas is an atlas of spatio-temporal gene expression profiles obtained by RNA sequencing and exon-array performed on macrodissected tissues at different stages of human development. Neither of these atlases corresponds with our goal of determining the whole brain gene expression profile using a wide range of data to provide a normalized reference expression value for each human transcript, with the chance of integrating them even if they come from multiple experiments carried out on different platforms (Lenzi et al., 2011).

The availability of the systematic and detailed expression maps presented here for several human tissues represented an excellent occasion to investigate, among many features of the transcriptome, the suitability of individual genes as the best ones for selection as reference genes in gene expression studies. This is because they fulfil criteria including a widely diffused, constant and high expression. The results consist of many known genes and uncharacterized loci, mostly EST clusters. The enhanced function is related to the known genes of each group. It is interesting to note that the non-brain tissue pool and the foetal brain share the same enriched function, i.e. structural constituent of ribosome (GO:0003735), a function associated with genes that result constitutively expressed in the cells although it is not associated with the same genes. Furthermore, the enriched function of the adult brain is unfolded protein binding (GO:0051082). We analyzed only these maps to verify if the housekeeping genes shared among the pool of 53 tissues are also present in the whole adult and foetal brain.

Finally, it would be interesting to investigate the role of certain ncRNAs emerging from our analysis when the corresponding probes were present on the experimental platform.

Our maps can provide a useful and ready reference benchmark to test hypotheses about localized gene expression levels of human transcripts in the brain and in three brain subregions such as the cerebellum, the cerbral cortex, and the hippocampus, regions severely affected in ID. These data could also contribute to a better understanding on a regional (chromosomal) basis of the chr21 gene expression (Strippoli et al., 2013). Its over-expression in trisomy 21 is associated with the most common form of constitutional ID.

The transcriptome maps can easily be extended to many other tissues and pathological conditions to obtain a quantitative dissection of regional gene expression levels within a certain tissue, or of differential expression between two biological conditions.

5.1.5 Whole heart transcriptome map

The quantitative transcriptome reference map of the whole normal human heart provided a reference typical value of expression for each the 43,360 known, mapped and 22,976 uncharacterized (unmapped) transcripts.

The heart transcriptome map showed the highest expression value of a segment on chr1 (1p36) and includes the over-expressed known genes *NPPA*, *NPPB* and *MFN2* encoding respectively for natriuretic peptide A, natriuretic peptide B and mitofusin 2 (Table 11). NPPA and NPPB belongs to the natriuretic peptide family and are synthesized as precursor proteins which undergo intracellular modification (cleavage events) to produce cardiac hormones involved in the cardiovascular homeostasis. Both are expressed in cardiac tissue and are used as diagnostic tools in cardiovascular disease (de Lemos et al., 2003; Song et al., 2005). MFN2 is a mitochondrial membrane protein, whose expression levels are used as a diagnostic tool for severe cardiovascular diseases as vascular proliferative disorders (Sorianello et al., 2012). Remarkably, genes encoding products with very different functions (endocrine and mytochondrial, respectively) result as belonging to the same cluster expressed at very high levels in the heart. The second over-expressed segment is on chr14 (14q11.2) (Table 11) and includes the over-expressed known genes: *PSMB5*, encoding for subunit beta 5 of the proteasome 20S, *MYH6*, encoding for myosin, heavy chain 6, cardiac muscle, alpha, *MYH7*, encoding for myosin, heavy chain 7, cardiac muscle, beta. It is well known that the *MYH6* map ~4kb downstream of *MYH7* (Matsuoka et al., 1989) and their expression and relative abundance of the proteins are correlated to the contractile velocity of cardiac muscle (Nakao et al., 1997). Both have been previously studied for their involvement in heart disease (Posch et al., 2011), with the conclusion that they are involved in different forms of congenital heart disease, both in association with mutation in the cardiac muscle actin gene (*ACTC1*) (Budde et al., 2007; Monserrat et al., 2007; Granados-Riveron et al., 2010).

At the single gene level, the over-expression of genes encoding for proteins involved in typical cardiac tissue structure and function emerge (Table 12). Cardiac muscle, although structurally and functionally similar to skeletal muscle, has some special features: a larger diameter of cardiac fibres necessary for the relatively high concentrations of stored cytosolic Ca2+, independent of extracellular supply; due to specialized groups of cells called pacemakers,

cardiac muscle is able to initiate its own contraction and rhythm without the requirement of external stimulation from motor neurons in order to contract; finally, myocardial cells are short, branched and interconnected with adjacent cells by gap junctions or electrical synapses (Davison et al., 2000). All these features enable a large mass of cardiac myocytes to contract autonomously as single cell block. The cardiac muscle fibres have a high aerobic oxidative capacity, and they consist of a great number of mitochondria and aerobic respiratory enzymes (Davison et al., 2000). The high demand of oxygen is also reflected in high levels of myoglobin and a richer capillary supply (Gray and Standring, 2005). Recent developments in genetic investigations and molecular analysis have greatly increased our understanding of the mechanisms involved during the process of muscle contraction. These mechanisms are well represented in the quantitative map of the heart transcriptome that we have obtained in that genes expressed at the highest levels are implied in the key structural and functional cardiac pathways, both considering the heart in itself as well as in comparison with non-cardiac tissues. This also suggests that other genes with less characterized roles in the heart but with a relevant expression in this organ may have critical functions for cardiac molecular physiopathology. In particular, the known genes with the highest levels of expression are: *MYL2* (expression value=10,216.56), *MYH7* (expression value=8,725.76) and *MB* (expression value=8,283.87) (Table 12), clearly associated to the contractile apparatus and the oxygen supply to it. They are followed by other genes involved in the main function of cardiac cells as *TCAP* (expression value=6,352.89), encoding for the protein involved in the titin linkage to Z-disc of sarcomer, *NPPA* (expression value=6,332.35), cited above, the gene for the cardiac muscle actin, *ACTC1* (expression value=5,779.15), and genes encoding for the three closely associated proteins of the troponin complex, i.e. troponin C (calcium binding), *TNNC1* (expression value=5,550.96), troponin I (inhibition of tropomyosin binding), *TNNI3* (expression value=4,648.05), and troponin T (tropomyosin binding), *TNNT2* (expression value=3,947.73) (Table III in the Data Supplement, it is available at: http://apollo11.isto.unibo.it/heart, in the event of acceptance of the manuscript it will be available at: http://apollo11.isto.unibo.it/suppl/it will be available at: http://apollo11.isto.unibo.it/suppl/). They encode for troponin proteins exclusively expressed in the cardiac muscle tissue distributed along the length of the thin filament and together with the tromyosin, *TPM1* (gene expression value=1,778.49) (Table III in the Data Supplement, it is available at: http://apollo11.isto.unibo.it/heart, in the event of acceptance of the manuscript it will be available at: http://apollo11.isto.unibo.it/suppl/) undertakes the control of initiation, inhibition and speed of contraction of skeletal and cardiac muscle (Davison et al., 2000). *TPM1* is one of the four known tropomyosin genes predominantly expressed in the heart (Davison et al., 2000), unlike the other isoforms (*TPM2*, *TPM3* and *TPM4*) which have a lower expression

level in cardiac tissue (Table III in the Data Supplement, it is available at: http://apollo11.isto.unibo.it/heart, in the event of acceptance of the manuscript it will be available at: http://apollo11.isto.unibo.it/suppl/), as was already known (Davison et al., 2000). A high level of expression was also observed for *TTN* (expression value=1,794.78), which completes the list of the genes encoding the main sarcomeric proteins. Being closely associated with the myosin thick filament, it is elastic and changes length as the sarcomer contracts and relaxes (Alberts et al., 2007). This quantitative representation of gene expression in the human heart, obtained without any a priori specific assumption and fully coherent with established biological knowledge about cardiac structure/function, seems to visualize at the molecular level the classical description of the basic histology of cardiac tissue.

Due to the fact that the heart is one of the organs principally involved in the typical symptoms of DS (Strippoli et al., 2013), particular attention was paid to the expression in the transcriptome map of chr21 genes, which were actually revealed to be over-expressed in transcriptome maps generated by TRAM when comparing diploid and trisomic blood cells (Pelleri et al. 2014). Among the first five chr21 genes, *ATP5J* and *ATP5O*, encoding for mitochondrial protein involved in the cell respiratory chain, are over-expressed, followed by *SOD1* (Table 12).

It is interesting to check if our transriptome model is coherent at a quantitative level with the ratio of gene products found at precise relative amounts in the heart. We will illustrate two typical examples: sarcomere structural proteins and proteins involved in the production of energy.

The sarcomere is a unique example in biology of a sub-molecular structure that has exceptional stability and organisation due to hundreds of molecules gathered among them. It is limited by two Z-discs, and consists of thin and thick filaments. The thin and thick filaments are the most important elements for the execution of the contraction cycle. The thick filaments consist mainly of myosin and myosin binding protein as MyBP-C. The thin filaments consist of cardiac actin, α-tropomyosin and C-, I- and T- troponins. The Z-disc, which separates myofibrils in sarcomeres, is the site that binds the thin filaments through alpha-actinin and the beta-subunits of CapZ protein and the giant protein, titin, through the T-cap (Sarantitis et al., 2012). Some of these proteins are described to interact among them with an expectable, if not accurate, stoichiometric ratio (Davison et al., 2000; Sarantitis et al., 2012). For instance, the C-, I- and T-troponins associate with each other to form the troponin complex in a stoichiometry of 1:1:1. We observed if there was a correlation between this stoichiometry and our whole normal human heart transcriptome map gene expression level among the specific cardiac troponin (*TNNC1*, *TNNI3* and *TNNT2*), and we found a ratio of 1.19:1 between *TNNC1* (expression

value=5,550.97) and *TNNI3* (expression value=4,648.05) genes, 1.18:1 between *TNNI3* and *TNNT2* (expression value=3,947.73), and 1.41:1 between *TNNC1* and *TNNT2*, suggesting that in this case, as it is often supposed, the main regulatory mechanisms controlling the amount of the protein are localized at the very initial step of gene expression. These observations confirm similar findings such as the coexpression at high and similar levels of hemoglobin A and B genes (*HBA* and *HBB*) in the blood (Piovesan et al., 2013).

Similarly, subunits ATP5J and ATP5O are known to be part of the same respiratory chain complex being present in equivalent amount. At the mRNA level, our map assigns the expression values to the relative encoding genes of 2,254.25 and 2,205.50, respectively, thus showing a 1.02:1 ratio fully consistent with the stoichiometry of their assembly in the mitochondrial ATP synthase complex (Kane and Van Eyk 2009).

However, the correspondence between transcripts and protein levels may not be granted as a rule, due to the mechanisms of post-transcriptional regulation of each gene (Denti et al., 2013) and also to the fact that proteins may belong to different omo- or etero-polymeric complexes with different stoichiometry.

To highlight the heart-specific gene expression profile, a comparison pool composed of 629 datasets derived from 53 different human tissues or organs has been accurately assembled.

The differential transcriptome map obtained by the comparison of the whole heart vs. the pool of tissues minus heart transcriptome map showed the over-expression of a segment on chr1 (Table 13) including the known genes for *ACTN2* and *RYR2*: the first is known to help the anchoring of the myofibrillar actin filaments to the Z disc of muscle cells (skeletal, cardiac) (Ribeiro et al., 2014) and the second is involved in the cardiac muscle excitation-contraction coupling, supplying calcium to the cardiac tissue (Liu et al., 2015b). Another segment on chr1 is over-expressed (Table 13), as in the single transcriptome map of the organ, and it includes the known genes: *NPPA*, *NPPB*, *MFN2* and seems to be one of those chromatin domains with a typical expression pattern in the cardiac tissue.

If we observe the first five under-expressed genome segments (Table 13): the segment on chr14 (14q32.13) including the cluster of genes encoding for members of the serpin peptidase inhibitor (*SERPINA1*, *SERPINA11*, *SERPINA5*) (Heit et al., 2013), and the segment on chr 8 (8p23.1), including the cluster of genes encoding for defensin protein (*DEFA6*, *DEFA4*, *DEFA1*, *DEFA5*) (Chapnik et al., 2012) involved in the host defence's immune system, emerge. Each of them seems to be chromatin domain including genes which have no function typical of cardiac tissue.

At single gene level the over-expression of troponin I type 3 (*TNNI3*) gene followed by troponin T type 2 (*TNNT2*) gene emerges (Table 14). The first is the actin-binding subunit, the

second is the tropomyosin-binding subunit of the troponin complex of the cardiac muscle (Solaro et al., 2013). Increased activity of these genes is known to cause hypertrophic cardiomiopathy (HCM) (Solaro et al., 2013). Mutations in genes encoding for sarcomeric proteins, as myosin, troponin and tropomyosin, are inherited mutations which can result in serious heart disease and are the main cause of HCM (Davison et al., 2000; Alberts et al, 2007). The same thing was observed for the nine genes encoding for myosin chain which have an expression ratio >2 in the differential transcriptome map (data unshown, see Supplementary TRAM file, it is available at: http://apollo11.isto.unibo.it/heart, in the event of acceptance of the manuscript it will be available at: http://apollo11.isto.unibo.it/suppl/). Four of them, if they undergo a heterozygous or omozygous missense mutation, cause severe cardiac defects and diseases. They encode for: MYH6 myosin, heavy chain 6, cardiac muscle, alpha (Carniel et al., 2005; Ching et al., 2005; Holm et al., 2011), MYL3 myosin, light chain 3, alkali, ventricular, skeletal, slow (Caleshu et al., 2011), MYH7 myosin, heavy chain 7, cardiac muscle, beta (Kamisago et al., 2000; Klaassen et al., 2008; Armel and Leinwand, 2009; Das et al., 2014), and MYL2 myosin, light chain 2, regulatory, cardiac, slow (Grey et al., 2005).

The interesting finding that a differential expression level in the heart for a structural heart protein encoding gene is highly predictive of cardiomyopathy (hypertrophic or dilated) or other structural heart defects led us to formulate the more generalized hypothesis that members of a gene family with a cellular function common to many cell types (e.g., potassium channels) but expressed at high levels in a specific tissue in comparison with other tissues may be systematically associated to functional disease in that tissue when mutated. We show here that among 101 potassium channel encoding genes, those expressed at a ratio >1.5 in the heart/non-cardiac tissues differential transcriptome map are significantly associated to human arrythmias. Instead, at the bottom of the gene list ordered by decreasing ratio values we note channels under-expressed in the heart, but known to be over-expressed in the brain in similar brain/non-brain tissues ratio values that are associated to neurologic phenotypes (Table V in the Data Supplement, it is available at: http://apollo11.isto.unibo.it/heart, in the event of acceptance of the manuscript it will be available at: http://apollo11.isto.unibo.it/suppl/). Remarkably, this is the first demonstration to our knowledge that quantitative expression values may be systematically used as clues to anticipate that effects of mutations that will most likely affect a given human tissue, so that the map described here could support further searching for a heart disease gene. This also opens the way to more generalized studies aimed at establishing this type of correlation for all the human genes and tissues and all the phenotypes described in humans. A study of this type has recently been published for the mouse (Liao and Weng, 2015): by analysing mRNA expression data derived from different methods together with phenotype data of mouse mutants,

it was shown that despite the presence of widespread ectopic transcription, gene expression and mutant phenotypes remain associated, particularly for tissue-specific genes, genes expressed at high level or genes expressed at early developmental stages. Our generation of quantitative, reference standard transcriptome maps for different human organs and tissues (here, and REFS TRAM Brain/Hyppo/AMKL), comparable because they are normalized around the mean and by scaled quantiles (Lenzi et al., 2011), could greatly facilitate a similar task for all human genes and tissues, along with the availability of the OMIM database containing systematic data about mutant human phenotypes.

Among the first five over-expressed chr21 genes of the differential transcriptome map, the over-expression of the genes for the mitochondrial proteins ATP5O and ATP5J return (Ronn et al., 2009; Zhu et al., 2013), in support to the significant role of mitochondria in the essential function played by the cardiac tissue (Table 14).

The differential transcriptome map results confirm that the heart specific expression pattern is quantitatively centred on sarcomeric and mitochondrial genes, in full agreement with its structure and energy-dependent contractile functions, as well as on hormonal peptide encoding genes, reflecting its well-known endocrine function (Ogawa and de Bold, 2014).

To prove the reliability of the heart transcriptome map generated by TRAM software, we performed an experimental validation of the obtained data. The significant correlation between the expected and observed data ($r=0.98$, p-value<0.0001) (Figure 7) allows us to consider all the other intervening values among the points we selected as *bona fide* values, thus for the first time providing a reliable and quantitative representation of the reference expression values for 43,360 mapped transcripts of the normal human heart.

Initial analysis of the biological meaning of these data clearly shows a relationship of our findings with what is already known about genes which have either a relevant function or are almost silenced in the heart.

The high significant correlation allowed us to consider each expression value as a reference expression value for a list of 43,360 known, mapped and 22,976 uncharacterized (unmapped) transcripts of the whole normal human heart, and confirmed that TRAM could be used to generate an atlas of the transcriptome of all kinds of biological conditions.

The availability of the systematic and detailed expression map presented here for the heart represented an excellent occasion to investigate, among many features of the transcriptome, the suitability of individual genes as the best stable reference genes for data normalization in gene expression studies. This is because they fulfill criteria including a widely diffused, constant and high expression (Casadei et al., 2011).

The best gene at behaving like a housekeeping in the heart is *MRPL51* (Table 17), encoding for mitochondrial ribosomal protein L51. The recurrent over-expression of mitochondrial protein is due to the typical structure of the cardiac cell, whose large volume proportion is occupied by mitochondria (Gray and Standring, 2005).

In our work about a quantitative reference global portrait of gene expression in the normal human heart we demonstrate the very high expression of the sarcomeric genes commonly described in the contractile system of the heart, we identify genes expressed especially in the heart vs. all other non-cardiac tissues, and we provide a list of genes with a great range of reference expression values, which could be used to infer a very high rate of information about the whole normal human heart.

To date researchers find much difficulty in analysing the results of gene expression studies due to difficulty in obtaining healthy cardiac muscle data and setting a control group to contrast the results. As a consequence, human studies of gene expression in cardiac muscle were limited. The great majority of studies are done in small groups of a particular gene or genes separately and exposed to a certain stimulus (Costa and Franco, 2015).

All produced data support the utility of the human heart map as a general standard reference and of the differential maps as tools to highlight differences for specific genes.
This kind of study could be applied to analyse the transcriptional profile of different biological conditions, for example of diseased myocardium, allowing the discovery of new biomarkers and the understanding of the mechanisms underlying pathophysiology (Liu et al., 2015a).

5.2 DS AMKL transcriptome

We also have presented here a comprehensive analysis of transcriptome in human DS AMKL cells. Integration of data from different sources, including data obtained from different Authors using a variety of platforms, was made possible by a recent approach described by us for creation and analysis of transcriptome maps (Ge et al., 2006). While most approaches are aimed to separate gene expression profiles related to the same biological source in subclasses, the TRAM tool provides means to integrate and summarize a pool of samples of the same biological origin leading to a global picture of gene expression for that condition. Moreover, TRAM identifies critical genomic regions and genes with significant differential expressions between two biological conditions.

Several Authors have determined gene expression profiles for DS or non-DS AMKL samples or have explicitly compared these two leukemic conditions. However, due to the rarity of the M7 subtype of leukemia and the need to limit the analysis to pediatric age because DS

AMKL occurs almost exclusively in children, these studies were typically limited to small group of samples. In addition, most platforms used in the microarray studies are affected by omissions or errors in mapping a certain percentage of probes to specific loci in the genome. In our analysis, the use of theversion of the TRAM software, TRAM 1.1, allowed us to map thousands of previously uncharacterized microarray probes and to avoid the errors in probe assignment to human loci often present in the data supplied by the manufacturer along with the platforms.

A systematic comparison of AMKL originated from trisomy 21 cells versus non-trisomy 21 cells should highlight specific mechanisms (Caldwell et al., 2013) related to the presence of an extra copy of chr21 in DS children developing AMKL. Moreover, we presented a comparison with normal MK cells that has never been performed in other analyses about AMKL. Our global quantitative models of the transcriptome in the AMKL cells could also be useful to test hypotheses for correlations between any parameter associated to the condition (e.g., specific mutations or phenotype aspects) and specific changes in gene expression.

Our results, obtained in an integrated and open setting without any a priori assumption, show several previously unidentified aspects regarding specificity of AMKL originated by trisomy 21 cells.

First, there are only a few genomic regions significantly over- or under-expressed when comparing DS versus non-DS AMKL samples. This finding suggests that transcriptome maps of these two conditions are similar while on the other hand allows to focus to a small set of regions that appears to be critical in order to differentiate these disease conditions. Relevant differences regarding genes were reported in Table 19 (genomic segments) and Table 20 (single genes), with potential implications for the identification of diagnostic or therapeutic targets. There are three main regions over-expressed in DS AMKL vs. non-DS AMKL (Table 19). The first one (15q21.2) contains *HDC* gene, whose mRNA is translated in the enzyme converting L-histidine to histamine produced by only a few cell types (Roela et al., 2007); *HDC* mRNA increase has been shown to be associated to basophilic rather than to MK differentiation of pluripotent hematopoietic cells (Malinge et al., 2012). These observations led to the discovery of a skewing toward a potential basophilic differentiation for DS AMKL not highlighted in the original works from which the data were derived. Supporting this hypothesis, *FCER1A* mRNA, encoding the alpha subunit of the high-affinity IgE receptor, the initiator of the allergic response and strongly typical of basophilic differentiation, is 6.3 times over-expressed (the 8th most over-expressed known gene) in DS vs. non-DS AMKL, reinforcing the notion that in DS AMKL but not in non-DS AMKL the leukemic dedifferentiation involved the possibility of redirection toward basophilic differentiation. Remarkably, in (Taub et al., 1999) is demonstrated by an electron microscopy analysis that AMKL blast cells from children with DS may contain basophil-like

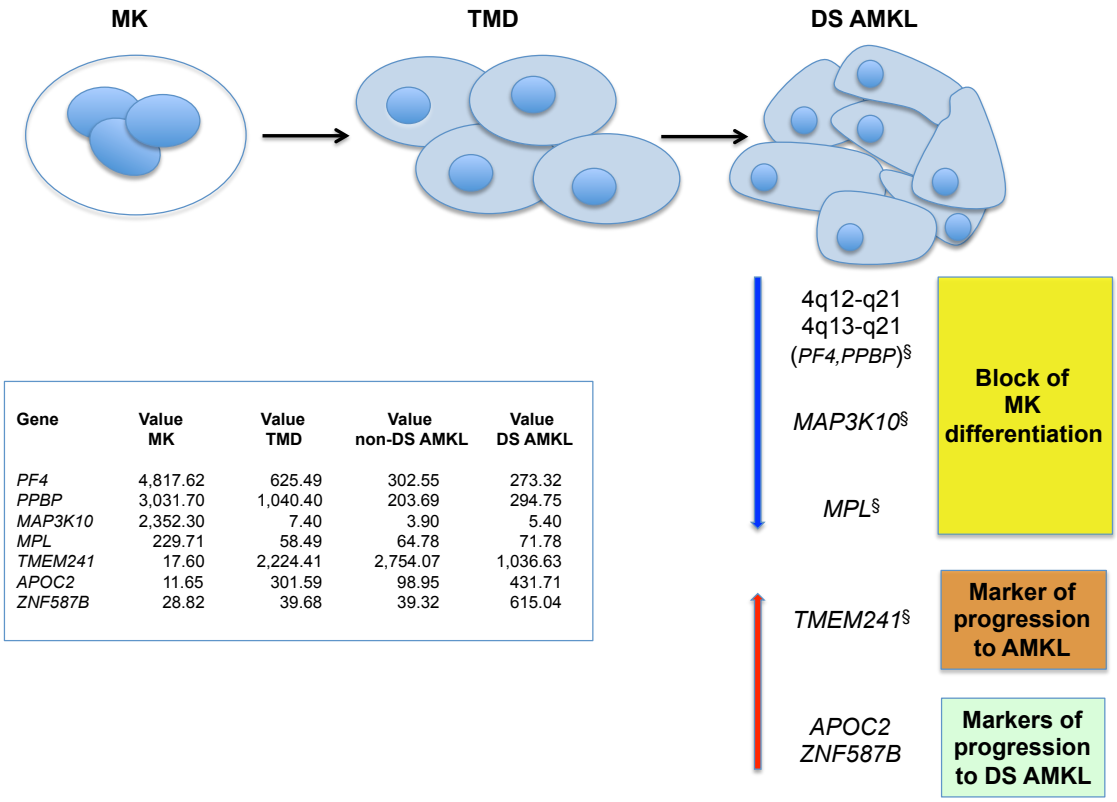
granules which were almost totally absent in blasts from children with non-DS AMKL or adults with AMKL, so that our data allows to the visualization of the molecular correlation at the level of the whole transcriptome of a morphological feature observed more than 20 years ago. Two other genomic regions are over-expressed in DS vs. non-DS AMKL blasts. The first is the region of glycophorins genes (*GYPE*, *GYPA*, *GYPB*) on chr4, erythroid surface markers (Boztug et al., 2013): this reinforces the concept of a disturbance of multilineage myeloid hematopoiesis in DS AMKL and has been observed by flow cytometry in the non-neoplastic hematopoiesis itself in trisomy 21 (Kaushansky, 2005). The other is the region of apolipoproteins genes (*APOC1*, *APOC2*, *APOE*) on chr19 that has been described as a signature of progression from TMD to DS AMKL in a gene expression profile (Lightfoot et al., 2004), it was not possible to include this in our analysis), further underlining its specificity for DS vs. non-DS MK blast cells.

When grouping expression values by chromosome, the chromosome with the greatest global RNA output was chr21 in both TMD and DS-AMKL vs. normal CB MK; we observed the same result in both DS and non-DS AMKL vs. normal MK comparisons. These data suggest that over-expression of chr21 genes is a key factor in AMKL development. In particular, *ADAMTS1*, encoding a protease known to inhibit angiogenesis (Caracausi et al., 2014) is the most over-expressed chr21 gene in DS vs. non-DS AMKL comparison. It is interesting to notice that this gene has been correlated to pediatric leukemias (ALL and DS-AML) (Strippoli et al., 2013; Melgarejo et al., 2010) but its exact role in leukemogenesis is still to be discovered. Our quantitative approach summarizing all values for each locus may clearly highlight mean expression ratio near to 1.5:1 for several chr21 genes when comparing DS AMKL vs. non-DS AMKL. This observation is consistent with the presence of an additional copy of the considered genes in trisomic cells. For example, *GABPA* expression presents a ratio close to 1.5:1 expected in trisomy of chr21, as well as other chr21 genes (*DYRK1A*, *SON*, *BACH1*) (Table 35). Even relatively small but significant differences (around 1.5-fold) in expression of numerous genes likely produce an aggregate effect, as observed in (Dy et al., 1999), where the same genes seem to be candidates to explain the impact of trisomy 21 in hematopoiesis abnormalities. For this reason it could be interesting to start from significant and robustus meta-analysis data to plan functional approaches in the future. Moreover, our data are consistent with the previous observation that *RUNX1*, *ERG* and *ETS2* oncogenes, although located on chr21, are not over-expressed in DS vs. non-DS AMKL (Bourquin et al., 2006; Klusmann et al., 2010b). It has also been recently demonstrated that they are not located on a chr21 duplicated minimal region in two cases of AML of M0 subtype (FAB classification) (Eguchi et al., 1992). As regard oncogenes, if chr21 oncogenes cited above appear not to be over-expressed in DS AMKL, *TRIB1* (chr8) may be a novel important oncogene for DS AMKL and its mutation is an earlier genetic event in

leukemogenesis (Nakahata and Okumura, 1994). In particular, it has been shown that a mutation of *TRIB1* (Nakahata and Okumura, 1994), a myeloid oncogene whose protein product is able to enhance ERK phosphorylation and to promote degradation of C/EBP family transcription factors, is a gain-of-function mutation remaining in leukocytes of the remission stage in which *GATA1* mutation disappeared. Our results show a mean value of 1.40 for the human *TRIB1* expression ratio between DS and non-DS AMKL samples, with a higher over-expression observed when comparing leukemic samples to normal MK (ratio 3.15 for DS AMKL and 2.26 for non-DS AMKL).

Although the expression of chr21 genes as a whole, or of some individual chr21 genes, may be coherent with the 3:2 gene expression ratio model in the comparison between trisomic and euploid cells, we note that discrepancies from this ideal model seen in our data for some of the comparisons we have made (Figures 8, 9 and 10) may be ascribed to the complexity of gene regulation in the aneuploid state (Maclean et al., 2012; Reynolds et al., 2010), to individual variability (Chang et al., 2010) as well as to the general dysregulation typical of the neoplastic state for which specific cell types we analyze is concerned.

Additional biologically relevant findings came from comparison of each type of megakaryoblastic leukemic condition with normal MK cells. Due to the role of the microenvironment in the hemopoiesis, including hemopoiesis in DS (Chou et al., 2012), it is expected that DS MK would present growth alterations due to trisomy 21 in both hemopoietic and microenvironment cells. From this point of view, DS MK cells would be the ideal control for the progression of a trisomic cell toward TMD and AMKL, however no gene expression profile dataset was available for this cell type. We propose here a biological model of the transcriptome depicting progressive changes from normal MK to TMD and then to DS AMKL, able to underline both shared and unique transcriptome map patterns for DS and non-DS variants of AMKL (Figure 19).



Gene	Value MK	Value TMD	Value non-DS AMKL	Value DS AMKL
PF4	4,817.62	625.49	302.55	273.32
PPBP	3,031.70	1,040.40	203.69	294.75
MAP3K10	2,352.30	7.40	3.90	5.40
MPL	229.71	58.49	64.78	71.78
TMEM241	17.60	2,224.41	2,754.07	1,036.63
APOC2	11.65	301.59	98.95	431.71
ZNF587B	28.82	39.68	39.32	615.04

Figure 19 Biological model of the transcriptome depicting progressive changes from MK to TMD then to DS AMKL. Downward pointing arrow indicates the repression of genes involved in MK differentiation; upward pointing arrow indicates the over-expression of potential molecular markers of progression to AMKL. Value: mean gene expression value normalized across all the pool samples. §Observed both in DS and non-DS AMKL.

Noteworthy, the genomic segment on chr4 known to contain a cluster of genes highly specific for MK differentiation (Canzonetta et al., 2012; Lenzi et al., 2011), was the highest significantly under-expressed segment in both DS and non-DS AMKL in comparison with normal MK. In particular, the more strongly under-expressed region 4q12-q21 contains a cluster of genes, including PF4 (encoding for the platelet factor 4, a main component of platelet alpha granules) (Antonarakis et al., 2004; Prandini et al., 2007; Yokoyama et al., 2012), and PPBP (encoding for beta-thromboglobulin) (Canzonetta et al., 2012), that are the most up-regulated transcripts in the megakaryocytic differentiation from CD34+ hematopoietic progenitors (Lenzi et al., 2011). This finding highlights a common final outcome of the block of MK cells differentiation in both DS and non-DS AMKL. It should be underlined that this result came from systematic ab initio analysis of more than 12,000 segments on the human genome including about 26,000 mapped loci, thus highlighting that this region critical for the MK differentiation is actually the more repressed in absolute when comparing transcriptome maps of AMKL (DS or non-DS) and normal MK cells.

In addition, the most under-expressed gene in TMD blasts when compared to normal MK cells is *MAP3K10*, encoding an activity of mitogen-activated protein kinase kinase kinase (MAPKKK). It is known that the mitogen-activated protein kinase (MAPK) pathway is involved in and is sufficient for megakaryocytic differentiation (Deutsch et al., 2005; Roberts and Izraeli, 2014). MAPK activity is present in several tens of human proteins, and we have identified the member *MAP3K10* as the critically repressed gene in the block of MK differentiation in the development of leukemia with MK features in that it appears down-regulated 300-fold in TMD cells and 500-fold in both DS and non-DS AMKL compared with normal MK samples. Finally, transcript for MPL, the receptor of thrombopoietin which is the primary regulator of normal thrombopoiesis (the formation of platelets) (Tunnacliffe et al., 1992; Gear and Camerini, 2003; Antonarakis et al., 2004), is decreased by ~70% in either TMD, DS and non-DS AMKL cells vs. normal MK cells.

An exceedingly high over-expression of the gene located on chromosome 18 and encoding for the uncharacterized membrane protein TMEM241 has been found in both DS (59-fold) and non-DS (156-fold) AMKL cells vs. normal MK. Although this probe was not present in all considered experimental platforms, its extreme differential expression makes it a candidate for further studies as a marker of progression from normal MK to AMKL blasts, also due to its 126-fold over-expression in TMD vs. MK cells.

Moreover, we identified several signatures of progression specifically to DS-AMKL. Remarkably, segments and genes up- or down-regulated in TMD in comparison with normal MK cells were highly similar to those specifically found in DS AMKL, underlining striking similarities between TMD and DS AMKL at the level of the whole transcriptome (already noted in (McElwaine et al., 2004), in their smaller set). On the other hand, a direct comparison between TMD and DS AMKL shows specific potential markers of progression to DS AMKL. As cited above, apolipoproteins genes (*APOC1*, *APOC2*, *APOE*) have been described as a signature of progression from TMD to DS AMKL (Lightfoot et al., 2004) and it is interesting to notice that *APOC2* is among the 20 most expressed genes in the comparison between TMD and MK (25.88-fold increase), showing a progressive increase of expression from normal MK to TMD and then to DS AMKL. In our analysis, *ZNF587B* appears to be the most discriminant marker between TMD and DS AMKL. Again, this observation offers the opportunity to integrate and discuss single genes and pathways previously described as abnormally expressed in DS or non-DS AMKL (Table 35). For example, the *PRAME* gene, encoding for a tumor antigen (McElwaine et al., 2004) was identified as a specific marker for DS AMKL blasts (n=7), with no expression in TMD (n=9). While our meta-analysis on *PRAME* expression data points (36 for DS AMKL and 11 for TMD) confirmed a clear over-expression of *PRAME* in DS AMKL (4.2-fold increase,

(Table 35)), it was not the most discriminant marker, that was exactly *ZNF587B*, while *PRAME* was the 33rd out of 25,955 transcripts ordered by decreasing DS AMKL vs. TMD expression ratio (Additional file 8, it is available at: http://apollo11.isto.unibo.it/suppl/2014_Pelleri/).

Finally, since leukemias in infants or young children originate from fetal hematopoietic cells (Klusmann et al., 2010a; Klusmann et al., 2010b; Chou et al., 2008; Tunstall-Pedoe et al., 2008) and the progenitors (fetal/neonatal MKP) are present in the cord blood (Olson et al., 1992; Liu and Sola-Visner, 2011), comparisons with CB MK cells have been also performed. Data from DS and non-DS AMKL vs. CB MK comparisons confirmed the repression of the clusters of genes expressed in MK. The over-expression of a region with collagen genes emerge both in DS and non-DS AMKL as well as that of the single gene *PTPRO* (Table 26 and 28), encoding for a tyrosine phosphatase receptor known to be involved in megakaryocytopoiesis and whose mRNA targeting by antisense oligonucleotides results in inhibited MK progenitor proliferation (Slungaard, 2005). On the other hand, the difference between DS and non-DS AMKL vs. CB MK is shown by the over-expression of the two olfactive receptor genes *OR10A5* and *OR10A4* (Table 26 and 28) only in non-DS derived cells.

The analysis of enrichment in specific gene functions using the tool FuncAssociate, for the 100 most over- or under-expressed genes in the comparison of DS vs. non-DS AMKL and of both of them vs. MK cells gave no significant results, other than a significant enrichment in genes involved in sequence-specific DNA-binding in non-DS AMKL vs. MK cells (data not shown), highlighting the relevance of remodeling the transcription factor network in leukemia.

Our results provide a systematic meta-analysis using any available gene expression profile dataset related to AMKL in pediatric age. These allow to identify more general trends and to produce a highly coherent view of the transcriptome depicting progressive changes from MK to TMD and then to DS AMKL. We believe that the originality of our results is due to several concurrent original features of the TRAM 1.1 platform. Advantages and relevant differences are: integration of the largest possible number of samples; integrated analysis of the largest possible number of genes (the integration of different platforms led us to assess expression ratio for about 26,000 loci, quantitating almost 4,000 genes in addition to the widely used platform U133A when used alone); absence of a priori filtering (in several works, this led to actual analysis of less than 50% of the genes present on the experimental platform); characterization at regional/map level in the study of gene expression (usually absent in the works from which data were obtained), relevant with regard to the study of an aneuploidy such as trisomy 21.

These results provide a new integrated model of the whole human transcriptome in DS and non DS AMKL, TMD and normal human MK cells, providing hints about pathophysiology of AMKL and also being useful to highlight possible clinical markers.

5.3 HR-DSCR on human chromosome 21 and gene hunting

A computational and molecular biology analysis of the HR-DSCR (Pelleri et al., 2016) was performed. This region of 34-kb was identified after a systematic reanalysis and comparison of all known and reported cases of PT21, enriched with analysis of Hsa21 CNVs in healthy subjects. Its duplication is shared by all the subjects with DS and is not present in all the subjects without DS. The main result discussed in the work of Pelleri and coll. (2016) is that, combining and comparing data which allow inclusion and exclusion of delimited Hsa21 stretches in the critical region for DS basic features, a very restricted region may be identified on Hsa21 within the traditional DSCR, confirming and further narrowing it to an unprecedented small size (0.07% of Hsa21) that is expected to contain a very small number of genetic determinants able to exert a major influence on the manifestation of DS basic features. Remarkably, analysis of gain (duplication) CNVs reported in 21q22 fully confirmed the peculiarity of this trait as it appears to be selectively spared by duplication in healthy subjects and it is included in a wider zone presenting with the minimal number of gain CNVs among all proximal and distal surrounding zones (Pelleri et al., 2016).

As a consequence of the identification of an HR-DSCR, a relevant causative role in DS for genes located outside of it should be rediscussed. There are no known genes located in HR-DSCR, while *KCNJ6* and *DSCR4* are the adjacent characterized genes, proximally and distally, respectively. Lack of support for the synergistic roles of *RCAN1*/*DSCR1* and *DYRK1A*, or *APP*, as main contributors to many DS phenotypes had already been underlined (Korbel et al., 2009). While a relevant causative role of *DYRK1A* for DS has been widely discussed (de la Torre and Dierrsen, 2012), and while it certainly affects the development and function of the nervous system, evidence that it is located within the HR-DSCR were not found, although it lies close to it (Pelleri et al., 2016). Interestingly, a subject reported by Cetin and coll. (2012) manifests DS in the absence of duplication of *DYRK1A* as shown by both FISH and Array CGH.

The definition of the HR-DSCR allows to exclude a critical role for DS as such for other known genes whose role in DS has often been discussed, although they may be involved in individual, non-constant DS phenotypes, such as *APP*, *SOD1*, *KCNJ6*, *ETS2* and *DSCAM* (Deitz et al., 2011).

Despite the predictions defining the HR-DSCR as intergenic, our research was focused on this region because no method of gene prediction is 100% reliable. This is demonstrated by the fact that new genes are continually described because the detailed analysis of a sequence of genomic DNA, computational biology analysis alone is not sufficient and must always be flanked by molecular biology analysis. Indeed, after the completion of the sequence of human

chromosome 21 and the resulting systematic cataloging of all active genes contained in it (Hattori et al., 2000), the large *CYYR1* gene coding for protein was identified (Vitale et al., 2002), not noticed in the previous analysis. The transcript has been identified thanks to partial homology with a single EST sequence of pig that, after analysis of molecular biology (RT-PCR, Northern Blot), has led to the description of a large locus of 107 Kb with final production of a short mature RNA encoding for a small protein of 150 amino acids containing a specific motif "CSSYAY" conserved in vertebrates. Therefore, it can be said that the catalog of human genes is constantly evolving and does not exclude the possibility of tracing novel uncharacterized transcripts.

Preliminary computational analysis of HR-DSCR nucleotide sequence shows that this region does not contain any known functional locus, that no undescribed protein-encoded gene is likely to lie in it and that good candidate microRNAs (miRNAs) sequences are present within it (preliminary data not shown). A miRNA might represent an excellent candidate for being a single major contributor to DS because of its pleiotropy of action. miRNA *MIR155* overexpression has been associated to DS through nexin 27 (Wang et al., 2013); however, no discussion was presented about its chromosomal localization proximally on 21q21.3 at 25.5 Mb, which is far from any described DSCR.

Our preliminary analysis with the TRAM program (Lenzi et al., 2011) identified 1,331 EST clusters (cluster) located on human chromosome 21 that confirm the presence of many partial sequences of RNA not yet characterized.

The systematic analysis of the region was performed also searching for the presence of transcribed sequences with regulatory function, particularly for microRNA sequences, taking into account the sequence of the putative miRNA1 for the high similarity to the known miRNA: hsa-mir 1273a.

We have developed methods of molecular biology in order to amplify the sequence of the putative miRNA1 (data unshown). These analyses, in particular the Northern Blot, led us to identify a sequence of about 13 Kb (in the heart and skeletal muscle) transcribed from the opposite strand with respect to that of the locus *DSCR4*, and the characterization of an active locus (*HRDSCR1*) transcribed bidirectionally.

The prediction of potential sequences for miRNAs in the considered region will allows us to characterize the *HRDSCR1* locus. Continuous interaction between methods of computational biology (*in silico*) and molecular biology (*in vitro*) is important to characterize novel transcripts, in particular in the human genome, as the complexity of organization shows the need for further work to make the catalog of human genes more complete.

Detailed analysis of the *HRDSCR1* locus has also led to the description of a new isoform of the known gene *DSCR4*. The gene was isolated from Nakamura et al. (1997), thanks to the finding of homology with an EST sequence and the subsequent cloning of the cDNA. The gene consists of 3 exons of 1095 bp and contains a long open reading frame that specifies 118 amino acids, whose sequence has no significant homology with any known protein. The gene is expressed in placental tissue and in adult skeletal muscle and heart tissues. The new isoform, for now conventionally known as *DSCR4.2*, shares exons 1 and 2 with the isoform 1, is devoid of exon 3, has at the 3' an intron and two additional exons, extends significantly the known locus *DSCR4* from 67,350 bp to 169,936 bp, conferring the status of transcribed region to more than the 100,000 bp previously considered intergenic. The locus *DSCR4* is only present in humans, and the wider DSCR region, responsible for DS, is fully conserved only in higher primates and in humans (Luke et al., 1995).

However, many studies have focused on DS using the mouse model in spite of its genome not having an ortholog region with significant homology to human DSCR (Takamatsu et al., 2002).

At this stage it is not possible to correlate the transcription profile observed in the locus to specific DS symptoms due to the lack of functional information in light of the observation that transcripts of the locus *HRDSCR1* do not appear to be expressed in the human brain, as it might be expected of genes involved in intellectual disability (the most constant symptom).

The bioinformatic and molecular investigation on the HR-DSCR allowed us to identify a new locus and to determine its complexity at the level of a region critical for Down syndrome, emphasizing the need to continue the analysis of the human genome, constantly integrating computational biology and molecular methods in order to complete, as far as possible, the list of human genes, accurately characterizing also the regions that are defined as intergenic.

In conclusion, further studies are needed to define properties and functions of transcripts considered, also in relation to the possible excision of these transcripts into functional products of small size, and their relationship with the DS phenotype. This study confirms the importance of a systematic analysis of apparently silent regions of the human genome because the variability and complexity of genomic organization of human genes does not, so far, consider the catalog of human active loci to be exhausted .

References

Alberts B, Johnson, A, Lewis, J, Raff, M, Roberts, K, Walter, P. 2007. The Cytoskeleton. In: Alberts B, Johnson, A, Lewis, J, Raff, M, Roberts, K, Walter, P, ed. Molecular Biology of the Cell, 5 ed. New York, USA: Garland Science, Taylor & Francis Group, pp 965-1052.

Alwine JC, Kemp DJ, Stark GR. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Biochemistry 74:5350-5354.

Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. 2015. OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic Acids Res 43:D789-798.

Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, Deutsch S. 2004. Chromosome 21 and down syndrome: from genomics to pathophysiology. Nat Rev Genet 5:725–738.

Antonarakis SE, Epstein CJ. 2006. The challenge of Down syndrome. Trends Mol Med 12:473-479.

Arisi GM. 2014. Nervous and immune systems signals and connections: Cytokines in hippocampus physiology and pathology. Epilepsy Behav 38C:43–47.

Armel TZ, Leinwand LA. 2009. Mutations in the beta-myosin rod cause myosin storage myopathy via multiple mechanisms. Proc Natl Acad Sci U S A 106:6291-6296.

Bao L, Loda M, Janmey PA, Stewart R, Anand-Apte B, Zetter BR. 1996. Thymosin beta 15: a novel regulator of tumor cell motility upregulated inmetastatic prostate cancer. NatMed 2:1322–1328.

Barrett T, Edgar R. 2006. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. Methods Enzymol 411:352–369.

Bartesaghi R, Guidi S, Ciani E. 2011. Is it possible to improve neurodevelopmental abnormalities in Down syndrome? Rev Neurosci 22:419–455.

Baumann MH, Wisniewski T, Levy E, Plant GT, Ghiso J. 1996. C-terminal fragments of alpha- and beta-tubulin form amyloid fibrils in vitro and associate with amyloid deposits of familial cerebral amyloid angiopathy, British type. Biochem Biophys Res Commun 219:238–242.

Bechstedt S, Brouhard GJ. 2012. Doublecortin recognizes the 13-protofilament microtubule cooperatively and tracks microtubule ends. Dev Cell 23:181–192.

Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, Xuan Z, Zhang MQ, Sedel F, Jourdren L, Coulpier F, et al. 2010. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. EMBO J 29:3082–3093.

Berriz GF, King OD, Bryant B, Sander C, Roth FP. 2003. Characterizing gene sets with FuncAssociate. Bioinformatics 19:2502–2504.

Bonsi L, Grossi A, Strippoli P, Tumietto F, Tonelli R, Vannucchi AM, Ronchi A, Ottolenghi S, Visconti G, Avanzi GC, et al. 1997. An erythroid and megakaryocytic common precursor cell line (B1647) expressing both c-mpl and erythropoietin receptor (Epo-R) proliferates and modifies globin chain synthesis in response to megakaryocyte growth and development factor (MGDF) but not to erythropoietin (Epo). Br J Haematol 98:549-559.

Boztug H, Schumich A, Pötschger U, Mühlegger N, Kolenova A, Reinhardt K, Dworzak M. 2013. Blast cell deficiency of CD11a as a marker of acute megakaryoblastic leukemia and transient myeloproliferative disease in children with and without Down syndrome. Cytometry B Clin Cytom 84:370–378.

Bourquin JP, Subramanian A, Langebrake C, Reinhardt D, Bernard O, Ballerini P, Baruchel A, Cavé H, Dastugue N, Hasle H, et al. 2006. Identification of distinct molecular phenotypes in acute megakaryoblastic leukemia by gene expression profiling. Proc Natl Acad Sci U S A 103:3339-3344.

Brooksbank C, Bergman MT, Apweiler R, Birney E and Thornton J. 2014. The European Bioinformatics Institute's data resources 2014. Nucl. Acids Res 42: D18-D25.

Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, Willard HF. 1992. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell 71:527–542.

Budde BS, Binner P, Waldmuller S, Hohne W, Blankenfeldt W, Hassfeld S, Bromsen J, Dermintzoglou A, Wieczorek M, May E, et al. 2007. Noncompaction of the ventricular myocardium is associated with a de novo mutation in the beta-myosin heavy chain gene. PLoS One 2:e1362.

Butte AJ, Dzau VJ, Glueck SB. 2001. Further defining housekeeping, or "maintenance," genes Focus on "A compendium of gene expression in normal human tissues". Physiol Genomics 7:95–96.

Caldwell JT, Edwards H, Dombkowski AA, Buck SA, Matherly LH, Ge Y, Taub JW. 2013. Over-expression of GATA1 confers resistance to chemotherapy in acute megakaryocytic Leukemia. PLoS One 8:e68601.

Caleshu C, Sakhuja R, Nussbaum RL, Schiller NB, Ursell PC, Eng C, De Marco T, McGlothlin D, Burchard EG, Rame JE. 2011. Furthering the link between the sarcomere and primary cardiomyopathies: restrictive cardiomyopathy associated with multiple mutations in genes previously associated with hypertrophic or dilated cardiomyopathy. Am J Med Genet A 155a:2229-2235.

Canzonetta C, Hoischen A, Giarin E, Basso G, Veltman JA, Nacheva E, Nizetic D, Groet J. 2012. Amplified segment in the 'Down syndrome critical region' on chr21 shared between Down syndrome and euploid AML-M0 excludes RUNX1, ERG and ETS2. Br J Haematol 157:197–200.

Caracausi M, Vitale L, Pelleri MC, Piovesan A, Bruno S, Strippoli P. 2014. A quantitative transcriptome reference map of the normal human brain. Neurogenetics 15:267-287.

Carlesimo GA, Marotta L, Vicari S. 1997. Long-term memory in mental retardation: evidence for a specific impairment in subjects with Down's syndrome. Neuropsychologia 35:71-79.

Carniel E, Taylor MR, Sinagra G, Di Lenarda A, Ku L, Fain PR, Boucek MM, Cavanaugh J, Miocic S, Slavov D, et al. 2005. Alpha-myosin heavy chain: a sarcomeric gene associated with dilated and hypertrophic phenotypes of cardiomyopathy. Circulation 112:54-59.

Casadei R, Pelleri MC, Vitale L, Facchin F, Lenzi L, Canaider S, Strippoli P, Frabetti F. 2011. Identification of housekeeping genes suitable for gene expression analysis in the zebrafish. Gene Expr Patterns 11:271–276.

Cetin Z, Yakut S, Mihci E, Manguoglu AE, Berker S, Keser I, Luleci G. 2012. A patient with Down syndrome with a de novo derivative chromosome 21. Gene 507:159-64.

Chang HH, McGeachie M, Alterovitz G, Ramoni MF. 2010. Mapping transcription mechanisms from multimodal genomic data. BMC Bioinformatics 11(Suppl 9):S2.

Chapnik N, Levit A, Niv MY, Froy O. 2012. Expression and structure/function relationships of human defensin 5. Appl Biochem Biotechnol 166:1703-1710.

Chen J, Bardes EE, Aronow BJ, Jegga AG. 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res 37:W305–W311.

Ching YH, Ghosh TK, Cross SJ, Packham EA, Honeyman L, Loughna S, Robinson TE, Dearlove AM, Ribas G, Bonser AJ, et al. 2005. Mutation in myosin heavy chain 6 causes atrial septal defect. Nat Genet 37:423-428.

Chou ST, Opalinska JB, Yao Y, Fernandes MA, Kalota A, Brooks JS, Choi JK, Gewirtz AM, Danet-Desnoyers GA, Nemiroff RL, Weiss MJ. 2008. Trisomy 21 enhances human fetal erythro-megakaryocytic development. Blood 112:4503-4506.

Chou ST, Byrska-Bishop M, Tober JM, Yao Y, Vandorn D, Opalinska JB, Mills JA, Choi JK, Speck NA, Gadue P, et al. 2012. Trisomy 21-associated defects in human primitive hematopoiesis revealed through induced pluripotent stem cells. Proc Natl Acad Sci U S A109:17573–17578.

Christa L, Carnot F, Simon MT, Levavasseur F, Stinnakre MG, Lasserre C, Thepot D, Clement B, Devinoy E, Brechot C. 1996. HIP/PAP is an adhesive protein expressed in hepatocarcinoma, normal Paneth, and pancreatic cells. Am J Physiol 271:G993–G1002.

Costa A, Franco OL. 2015. Insights into RNA transcriptome profiling of cardiac tissue in obesity and hypertension conditions. J Cell Physiol 230:959-968.

Costa AC, Scott-McKean JJ. 2013. Prospects for improving brain function in individuals with Down syndrome. CNS Drugs 27:679–702.

Czech T, Yang JW, Csaszar E, Kappler J, Baumgartner C, Lubec G. 2004. Reduction of hippocampal collapsin response mediated protein-2 in patients with mesial temporal lobe epilepsy. Neurochem Res 29:2189–2196.

Das KJ, Ingles J, Bagnall RD, Semsarian C. 2014. Determining pathogenicity of genetic variants in hypertrophic cardiomyopathy: importance of periodic reassessment. Genet Med 16:286-293.

Davis LG, KuehlWM, Battey JF. 1994. Basic methods in molecular biology. Appleton & Lange, Norwalk.

Davison FD, D'Cruz LG, McKenna WJ. 2000. Molecular motors in the heart. Essays Biochem 35:145-158.

de la Torre R, Dierssen M. 2012. Therapeutic approaches in the improvement of cognitive performance in Down syndrome: past, present, and future. Prog Brain Res 197:1-14.

de Lemos JA, McGuire DK, Drazner MH. 2003. B-type natriuretic peptide in cardiovascular disease. Lancet 362:316-322.

Deitz SL, Blazek JD, Solzak JP, Roper RJ. 2011. Down syndrome: A complex and interactive genetic disorder. In: Dey S, ed. Genetics and Etiology of Down Syndrome. Rijeka, CR: InTech, pp 65-96.

Delabar JM, Theophile D, Rahmani Z, Chettouh Z, Blouin JL, Prieur M, Noel B, Sinet PM. 1993. Molecular mapping of twenty-four features of Down syndrome on chromosome 21. Eur J Hum Genet 1:114-124.

Denti MA, Viero G, Provenzani A, Quattrone A, Macchi P. 2013. mRNA fate: Life and death of the mRNA in the cytoplasm. RNA Biol 10:360-366.

Deutsch S, Lyle R, Dermitzakis ET, Attar H, Subrahmanyan L, Gehrig C, Parand L, Gagnebin M, Rougemont J, Jongeneel CV, Antonarakis SE. 2005. Gene expression variation and expression quantitative trait mapping of human chromosome 21 genes. Hum Mol Genet 14:3741–3749.

Dierssen M, Herault Y, Estivill X. 2009. Aneuploidy: from a physiological mechanism of variance to Down syndrome. Physiol Rev 89:887-920

Dmitrieva J, Chen C, Greenberger E, Ogunseitan O, Ding YC. 2011. Gender-specific expression of the DRD4 gene on adolescent delinquency, anger and thrill seeking. Soc Cogn Affect Neurosci 6:82–89.

Dy M, Pacilio M, Arnould A, Machavoine F, Mayeux P, Hermine O, Bodger M, Schneider E. 1999. Modulation of histidine decarboxylase activity and cytokine synthesis in human leukemic cell lines: relationship with basophilic and/or megakaryocytic differentiation. Exp Hematol 27:1295–1305.

D'Hulst C, De Geest N, Reeve SP, Van Dam D, De Deyn PP, Hassan BA, Kooy RF. 2006. Decreased expression of the GABAA receptor in fragile X syndrome. Brain Res 1121:238–245.

Eguchi M, Ozawa T, Sakakibara H, Sugita K, Iwama Y, Furukawa T. 1992. Ultrastructural and ultracytochemical differences between megakaryoblastic leukemia in children and adults. Analysis of 49 patients. Cancer 70:451–458.

Emig D, Salomonis N, Baumbach J, Lengauer T, Conklin BR, Albrecht M. 2010. AltAnalyze and DomainGraph: Analyzing and visualizing exon expression data. Nucleic Acids Res 38:W755–W762.

Enard W, Khaitovich P, Klose J, Zöllner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, et al. 2002. Intra- and interspecific variation in primate gene expression patterns. Science 296:340-343.

ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489:57-74. Engels WR. 1993. Contributing software to the internet: the Amplify program. Trends Biochem Sci 18:448-450.

Epstein CJ. 1989. Down syndrome, trisomy 21. In: Scriver CR, Beaudet AL, Sly WS, Valle D. Metabolic Basis of Inherited Disease. McGraw-Hill, New York, NY, pp 91-326.

Ervin JF, Heinzen EL, Cronin KD, Goldstein D, Szymanski MH, Burke JR,Welsh-Bohmer KA, Hulette CM. 2007. Postmortem delay hasminimal effect on brain RNA integrity. J Neuropathol Exp Neurol 66:1093–1099.

Fallin MD, Lasseter VK, Liu Y, Avramopoulos D, McGrath J, Wolyniec PS, Nestadt G, Liang KY, Chen PL, Valle D, Pulver AE. 2011. Linkage and association on 8p21.2-p21.1 in schizophrenia. Am J Med Genet B Neuropsychiatr Genet 156:188–197.

Fanarraga ML, Villegas JC, Carranza G, Casta~no R, Zabala JC. 2009. Tubulin cofactor B regulates microtubule densities during microglia transition to the reactive states. Exp Cell Res 315:535–541.

Fatemi SH, Reutiman TJ, Folsom TD, Thuras PD. 2009. GABA(A) receptor downregulation in brains of subjects with autism. J Autism Dev Disord 39:223–230.

Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, et al. 2009. The regulated retrotransposon transcriptome of mammalian cells. Nat Genet 41:563–571.

Felli N, Cianetti L, Pelosi E, Carè A, Liu CG, Calin GA, Rossi S, Peschle C, Marziali G, Giuliani A. 2010. Hematopoietic differentiation: a coordinated dynamical process towards attractor stable states. BMC Syst Biol 4:85.

Ferrari F, Bortoluzzi S, Coppe A, Basso D, Bicciato S, Zini R, Gemelli C, Danieli GA, Ferrari S. 2007. Genomic expression during human myelopoiesis. BMC Genomics 8:264–283.

Fuhrken PG, Chen C, Miller WM, Papoutsakis ET. 2007 Comparative, genome-scale transcriptional analysis of CHRF-288-11 and primary human megakaryocytic cell cultures provides novel insights into lineage-specific differentiation. Exp Hematol 35:476–489.

Fung SJ, Joshi D, Allen KM, Sivagnanasundaram S, Rothmond DA, Saunders R, Noble PL, Webster MJ, Weickert CS. 2011. Developmental patterns of doublecortin expression and white matter neuron density in the postnatal primate prefrontal cortex and schizophrenia. PLoS One 6:e25194.

Gardiner K, Costa AC. 2006. The proteins of human chromosome 21. Am J Med Genet C Semin Med Genet 142C(3):196-205.

Gardiner KJ. 2010. Molecular basis of pharmacotherapies for cognition in Down syndrome. Trends Pharmacol Sci 31:66-73.

Ge Y, Dombkowski AA, LaFiura KM, Tatman D, Yedidi RS, Stout ML, Buck SA, Massey G, Becton DL, Weinstein HJ, et al. 2006. Differential gene expression, GATA1 target genes, and the chemotherapy sensitivity of Down syndrome megakaryocytic leukemia. Blood 107:1570-1581.

Gear AR, Camerini D. 2003. Platelet chemokines and chemokine receptors: linking hemostasis, inflammation, and host defense. Microcirculation 10:335–350.

Giammona LM, Fuhrken PG, Papoutsakis ET, Miller WM. 2006. Nicotinamide (vitamin B3) increases the polyploidisation and proplatelet formation of cultured primary human megakaryocytes. Br J Haematol 135:554–566.

Granados-Riveron JT, Ghosh TK, Pope M, Bu'Lock F, Thornborough C, Eason J, Kirk EP, Fatkin D, Feneley MP, Harvey RP, et al. 2010. Alpha-cardiac myosin heavy chain (MYH6) mutations affecting myofibril formation are associated with congenital heart defects. Hum Mol Genet 19:4007-4016.

Gray H, Standring S. 2005. Heart and great vessels. In: Standring S, ed. Gray's Anatomy: The Anatomical Basis of Clinical Practice, 39 ed. Philadelphia: Churchill Livingstone, pp 995-1020.

Grey C, Mery A, Puceat M. 2005. Fine-tuning in Ca2+ homeostasis underlies progression of cardiomyopathy in myocytes derived from genetically modified embryonic stem cells. Hum Mol Genet 14:1367-1377.

Gu YM, Li SY, Qiu XS,Wang EH. 2008. Elevated thymosin beta15 expression is associated with progression andmetastasis of non-small cell lung cancer. APMIS 116:484–490.

Guo CJ, Douglas SD, Lai JP, Pleasure DE, Li Y, Williams M, Bannerman P, Song L, Ho WZ. 2003. Interleukin-1beta stimulates macrophage inflammatory protein-1alpha and -1beta expression in human neuronal cells (NT2-N). J Neurochem 84:997–1005.

Guo Y, Sheng Q, Li J, Ye F, Samuels DC, Shyr Y. 2013. Large scale comparison of gene expression levels by microarrays and RNAseq using TCGA data. PLoS One 8:e71462.

Hartman RJ, Rasmussen SA, Botto LD, Riehle-Colarusso T, Martin CL, Cragan JD, Shin M, Correa A. 2011. The contribution of chromosomal abnormalities to congenital heart defects: a population-based study. Pediatr. Cardiol 32: 1147–1157.

Hasle H, Clemmensen IH, Mikkelsen M. 2000. Risks of leukaemia and solid tumours in individuals with Down's syndrome. Lancet 355:165-169.

Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T, Park HS, Toyoda A, Ishii K, Totoki Y, Choi DK, et al. 2000. Chromosome 21 mapping and sequencing consortium. The DNA sequence of human chromosome 21. Nature 405:311-319.

Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, van de Lagemaat LN, Smith KA, Ebbert A, Riley ZL, et al. 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 489:391–399.

Haydar TF, Reeves RH. 2012. Trisomy 21 and early brain development. Trends Neurosci 35:81-91.

Hehr U, Uyanik G, Aigner L, Couillard-Despres S,Winkler J. 2007. DCX-related disorders. In: Pagon RA, Adam MP, Bird TD, Dolan CR, Fong CT, Smith RJH, Stephens K (eds) GeneReviews® [Internet]. University of Washington, Seattle, Seattle, 1993–2014.

Heit C, Jackson BC, McAndrews M, Wright MW, Thompson DC, Silverman GA, Nebert DW, Vasiliou V. 2013. Update of the human and mouse SERPIN gene superfamily. Hum Genomics 7:22.

Herrera R, Hubbell S, Decker S, Petruzzelli L. 1998. A role for the MEK/MAPK pathway in PMA-induced cell cycle arrest: modulation of megakaryocytic differentiation of K562 cells. Exp Cell Res 238:407–414.

Hickey F, Hickey E, Summar KL. 2012. Medical update for children with Down syndrome for the pediatrician and family practitioner. Adv Pediatr 59:137-517.

Hitzler JK, Zipursky A. 2005. Origins of leukaemia in children with Down syndrome. Nat Rev Cancer 5:11-20.

Hodges A, Strand AD, Aragaki AK, Kuhn A, Sengstag T, Hughes G, Elliston LA, Hartog C, Goldstein DR, Thu D, et al. 2006. Regional and cellular gene expression changes in human Huntington's disease brain. Hum Mol Genet 15:965-977.

Holm H, Gudbjartsson DF, Sulem P, Masson G, Helgadottir HT, Zanon C, Magnusson OT, Helgason A, Saemundsdottir J, Gylfason A, et al. 2011. A rare variant in MYH6 is associated with high risk of sick sinus syndrome. Nat Genet 43:316-320.

Iijima T,Miura E,Matsuda K, Kamekawa Y,WatanabeM, YuzakiM. 2007. Characterization of a transneuronal cytokine family Cbln–regulation of secretion by heteromeric assembly. Eur J Neurosci 25:1049–1057.

Iwamoto K, Kakiuchi C, Bundo M, Ikeda K, Kato T. 2004. Molecular characterization of bipolar disorder by comparing gene expression profiles of postmortem brains of major mental disorders. Mol Psychiatry 9:406-416.

Jäger D, Stockert E, Güre AO, Scanlan MJ, Karbach J, Jäger E, Knuth A, Old LJ, Chen YT. 2001. Identification of a tissuespecific putative transcription factor in breast tissue by serological screening of a breast cancer library. Cancer Res 61:2055–2061.

Kamisago M, Sharma SD, DePalma SR, Solomon S, Sharma P, McDonough B, Smoot L, Mullen MP, Woolf PK, Wigle ED, et al. 2000. Mutations in sarcomere protein genes as a cause of dilated cardiomyopathy. N Engl J Med 343:1688-1696.

Kane LA, Van Eyk JE. 2009. Post-translational modifications of ATP synthase in the heart: biology and function. J Bioenerg Biomembr 41:145-150.

Kaushansky K. 2005. The molecular mechanisms that control thrombopoiesis. J Clin Invest 115:3339–3347.

Kesslak JP, Nagata SF, Lott I, Nalcioglu O. 1994. Magnetic resonance imaging analysis of age-related changes in the brains of individuals with Down's syndrome. Neurology 44:1039-1045.

Khaitovich P, Muetzel B, She X, Lachmann M, Hellmann I, Dietzsch J, Steigele S, Do HH, Weiss G, Enard W et al. 2004. Regional patterns of gene expression in human and chimpanzee brains. Genome Res 14:1462-1473.

Khan I, Malinge S, Crispino J. 2011. Myeloid leukemia in Down syndrome. Crit Rev Oncog 16:25-36.

Kikushige Y, Shima T, Takayanagi S, Urata S, Miyamoto T, Iwasaki H, Takenaka K, Teshima T, Tanaka T, Inagaki Y, Akashi K. 2010. TIM-3 is a promising target to selectively kill acute myeloid leukemia stem cells. Cell Stem Cell 7:708–717.

Klaassen S, Probst S, Oechslin E, Gerull B, Krings G, Schuler P, Greutmann M, Hurlimann D, Yegitbasi M, Pons L, et al. 2008. Mutations in sarcomere protein genes in left ventricular noncompaction. Circulation 117:2893-2901.

Klusmann JH, Godinho FJ, Heitmann K, Maroz A, Koch ML, Reinhardt D, Orkin SH, Li Z. 2010a. Developmental stage-specific interplay of GATA1 and IGF signaling in fetal megakaryopoiesis and leukemogenesis. Genes Dev 24:1659-1672.

Klusmann JH, Li Z, Böhmer K, Maroz A, Koch ML, Emmrich S, Godinho FJ, Orkin SH, Reinhardt D. 2010b. miR-125b-2 is a potential oncomiR on human chromosome 21 in megakaryoblastic leukemia. Genes Dev 24:478-490.

Korbel JO, Tirosh-Wagner T, Urban AE, Chen XN, Kasowski M, Dai L, Grubert F, Erdman C, Gao MC, Lange K, et al. 2009. The genetic architecture of Down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies. Proc Natl Acad Sci U S A 106:12031-12036.

Korenberg JR, Kawashima H, Pulst SM, Ikeuchi T, Ogasawara N, Yamamoto K, Schonberg SA, West R, Allen L, Magenis E, et al. 1990. Molecular definition of a region of chromosome 21 that causes features of the Down syndrome phenotype. Am J Hum Genet 47:236-246.

Korenberg JR. 2009. Down syndrome: the crucible for treating genomic imbalance. Genet Med 11:617-619.

Kozomara A, Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42(Database issue):D68-73.

Kremerskothen J, Kindler S, Finger I, Veltel S, Barnekow A. 2006. Postsynaptic recruitment of Dendrin depends on both dendritic mRNA transport and synaptic anchoring. J Neurochem 96:1659–1666.

Kryger R, Fan L, Wilce PA, Jaquet V. 2012. MALAT-1, a non protein-coding RNA is upregulated in the cerebellum, hippocampus and brain stem of human alcoholics. Alcohol 46:629–634.

Kuhn RM, Karolchik D, Zweig AS,Wang T, Smith KE, Rosenbloom KR, Rhead B, Raney BJ, Pohl A, Pheasant M, et al. 2009. The UCSC Genome Browser Database: update 2009. Nucleic Acids Res 37:D755–D761.

Kumar RA, Pilz DT, Babatz TD, Cushion TD, Harvey K, Topf M, Yates L, Robb S, Uyanik G, Mancini GM, et al. 2010. TUBA1A mutations cause wide spectrum lissencephaly (smooth brain)

and suggest that multiple neuronal migration pathways converge on alpha tubulins. Hum Mol Genet 19:2817–2827.

Lejeune J, Gautier M, Turpin R. 1959. Etude des chromosomes somatiques de neuf enfants mongoliens. C R Hebd Seances Acad Sci 248:1721-1722.

Lejeune J. 1988. Research on pathogeny of mental retardation in trisomy 21. Working group on: "Aspects of the uses of genetic engineering". Commentarii Vol. III N° 31. Pontificia Academia Scientiarum, Rome, Italy.

Lenzi L, Frabetti F, Facchin F, Casadei R, Vitale L, Canaider S, Carinci P, Zannotti M, Strippoli P. 2006. UniGene Tabulator: A full parser for the UniGene format. Bioinformatics 22:2570–2571.

Lenzi L, Facchin F, Piva F, Giulietti M, Pelleri MC, Frabetti F, Vitale L, Casadei R, Canaider S, Bortoluzzi S, et al. 2011. TRAM (Transcriptome Mapper): database-driven creation and analysis of transcriptome maps from multiple sources. BMC Genomics 12:121.

Letourneau A, Antonarakis SE. 2012.Genomic determinants in the phenotypic variability of Down syndrome. Prog Brain Res 197:15-28.

Li HY, Li JF, Lu GW. 2003. Neurogranin: a brain-specific protein. Sheng Li Ke Xue Jin Zhan 34:111–115.

Liao BY, Weng MP. 2015. Unraveling the association between mRNA expressions and mutant phenotypes in a genome-wide assessment of mice. Proc Natl Acad Sci U S A 112:4707-4712.

Lightfoot J, Hitzler JK, Zipursky A, Albert M, Macgregor PF. 2004. Distinct gene signatures of transient and acute megakaryoblastic leukemia in Down syndrome. Leukemia 18:1617-1623.

Lima FA, Moreira-Filho CA, Ramos PL, Brentani H, Lima Lde A, Arrais M, Bento-de-Souza LC, Bento-de-Souza L, Duarte MI, Coutinho A, Carneiro-Sampaio M. 2011. Decreased AIRE expression and global thymic hypofunction in Down syndrome. J Immunol 187:3422–3430.

Liu Y, Morley M, Brandimarto J, Hannenhalli S, Hu Y, Ashley EA, Tang WH, Moravec CS, Margulies KB, Cappola TP, Li M. 2015a. RNA-Seq identifies novel myocardial gene expression signatures of heart failure. Genomics 105:83-89.

Liu Y, Sun B, Xiao Z, Wang R, Guo W, Zhang JZ, Mi T, Wang Y, Jones PP, Van Petegem F, Chen SR. 2015b. Roles of the NH2-terminal domains of cardiac ryanodine receptor in Ca2+ release activation and termination. J Biol Chem 290:7736-7746.

Liu ZJ, Sola-Visner M. 2011. Neonatal and adult megakaryopoiesis. Curr Opin Hematol 18:330-337.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25:402–408.

Lockhart DJ, Barlow C. 2001. Expressing what's on your mind: DNA arrays and the brain. Nat Rev Neurosci 2:63-68.

Lockstone HE, Harris LW, Swatton JE, Wayland MT, Holland AJ, Bahn S. 2007. Gene expression profiling in the adult Down syndrome brain. Genomics 90:647–660.

Lu J, Lian G, Zhou H, Esposito G, Steardo L, Delli-Bovi LC, Hecht JL, Lu QR, Sheen V. 2012. OLIG2 over-expression impairs proliferation of human Down syndrome neural progenitors. Hum Mol Genet 21:2330–2340.

Luke S, Gandhi S, Verma RS. 1995. Conservation of the Down syndrome critical region in humans and great apes. Gene 161:283-285.

Lyle R, Béna F, Gagos S, Gehrig C, Lopez G, Schinzel A, Lespinasse J, Bottani A, Dahoun S, Taine L, et al. 2009. Genotype-phenotype correlations in Down syndrome identified by array CGH in 30 cases of partial trisomy and partial monosomy chromosome 21. Eur J Hum Genet 17:454-466.

Maclean GA, Menne TF, Guo G, Sanchez DJ, Park IH, Daley GQ, Orkin SH. 2012. Altered hematopoiesis in trisomy 21 as revealed through in vitro differentiation of isogenic human pluripotent cells. Proc Natl Acad Sci U S A 2012, 109:17567–17572.

Malinge S, Bliss-Moreau M, Kirsammer G, Diebold L, Chlon T, Gurbuxani S, Crispino JD. 2012. Increased dosage of the chromosome 21 ortholog Dyrk1a promotes megakaryoblastic leukemia in a murine model of Down syndrome. J Clin Invest 122:948–962.

Malone JH, Oliver B. 2011. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol 9:34.

Martignetti JA, Brosius J. 1993. BC200 RNA: a neural RNA polymerase III product encoded by a monomeric Alu element. Proc Natl Acad Sci U S A 90:11563–11567.

Massey GV. 2005. Transient leukemia in newborns with Down syndrome. Pediatr Blood Cancer 44:29-32.

Matsuoka R, Yoshida MC, Kanda N, Kimura M, Ozasa H, Takao A. 1989. Human cardiac myosin heavy chain gene mapped within chromosome region 14q11.2----q13. Am J Med Genet 32:279-284.

McElwaine S, Mulligan C, Groet J, Spinelli M, Rinaldi A, Denyer G, Mensah A, Cavani S, Baldo C, Dagna-Bricarelli F, et al. 2004. Microarray transcript profiling distinguishes the transient from the acute type of megakaryoblastic leukaemia (M7) in Down's syndrome, revealing PRAME as a specific discriminating marker. Br J Haematol 125:729-742.

Mégarbané A, Ravel A, Mircher C, Sturtz F, Grattau Y, Rethoré MO, Delabar JM, Mobley WC. 2009. The 50th anniversary of the discovery of trisomy 21: the past, present, and future of research and treatment of Down syndrome. Genet Med 11:611-616.

Mégarbané A, Noguier F, Stora S, Manchon L, Mircher C, Bruno R, Dorison N, Pierrat F, Rethoré MO, Trentin B, et al. 2013. The intellectual disability of trisomy 21: differences in gene expression in a case series of patients with lower and higher IQ. Eur J Hum Genet 21:1253-1259.

Melemed AS, Ryder JW, Vik TA. 1997. Activation of the mitogen-activated protein kinase pathway is involved in and sufficient for megakaryocytic differentiation of CMK cells. Blood 90:3462–3470.

Melgarejo E, Medina MA, Sánchez-Jiménez F, Urdiales JL. 2010. Targeting of histamine producing cells by EGCG: a green dart against inflammation? J Physiol Biochem 66:265–270.

Melis D, Genesio R, Cappuccio G, MariaGinocchio V, Casa RD, Menna G, Buffardi S, Poggi V, Leszle A, Imperati F, et al. 2011. Mental retardation, congenital heart malformation, and myelodysplasia in a patient with a complex chromosomal rearrangement involving the critical region 21q22. Am J Med Genet A 155A:1697-1705.

Monserrat L, Hermida-Prieto M, Fernandez X, Rodriguez I, Dumont C, Cazon L, Cuesta MG, Gonzalez-Juanatey C, Peteiro J, Alvarez N, et a. 2007. Mutation in the alpha-cardiac actin gene associated with apical hypertrophic cardiomyopathy, left ventricular non-compaction, and septal defects. Eur Heart J 28:1953-1961.

Morris JK, Wald NJ, Watt HC. 1999. Fetal loss in Down syndrome pregnancies. Prenat Diagn 19:142-145.

Mus E, Hof PR, TiedgeH. 2007. Dendritic BC200 RNAin aging and in Alzheimer's disease. Proc Natl Acad Sci U S A 104:10679–10684.

Nadel L. 2003. Down's syndrome: a genetic disorder in biobehavioral perspective. Genes Brain Behav 2:156-166.

Naderi A, Teschendorff AE, Beigel J, Cariati M, Ellis IO, Brenton JD, Caldas C. 2007. BEX2 is over-expressed in a subset of primary breast cancers and mediates nerve growth factor/nuclear factor-kappaB inhibition of apoptosis in breast cancer cell lines. Cancer Res 67:6725–6736.

Nakahata T, Okumura N. 1994. Cell surface antigen expression in human erythroid progenitors: erythroid and megakaryocytic markers. Leuk Lymphoma 13:401–409.

Nakamura A, Hattori M, Sakaki Y. 1997. A novel gene isolated from human placenta located in Down syndrome critical region on chromosome 21. DNA Res 4:321-324.

Nakao K, Minobe W, Roden R, Bristow MR, Leinwand LA. 1997. Myosin heavy chain gene expression in human heart failure. J Clin Invest 100:2362-2370.

National Human Genome Research Institute (NHGRI) Online Education Kit: Understanding the Human Genome Project. Dynamic Timeline, 2013 http://www.genome.gov/25019887.

Nover-shtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME, Habib N, Yosef N, Chang CY, Shay T, et al. 2011. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell 144:296–309.

Nunez J. 1985. Microtubules and brain development: the effects of thyroid hormones. Neurochem Int 7:959–968.

Ogawa T, de Bold AJ. 2014. The heart as an endocrine organ. Endocr Connect 3:R31-44.

Okumura A, Hayashi M, Tsurui H, Yamakawa Y, Abe S, Kudo T, Suzuki R, Shimizu T, Shimojima K, Yamamoto T. 2013. Lissencephaly with marked ventricular dilation, agenesis of corpus callosum, and cerebellar hypoplasia caused by TUBA1A mutation. Brain Dev 35:274–279.

Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH. 2008. Functional organization of the transcriptome in human brain. Nat Neurosci 11:1271-1282.

Olson TA, Levine RF, Mazur EM, Wright DG, Salvado AJ. 1992. Megakaryocytes and megakaryocyte progenitors in human cord blood. Am J Pediatr Hematol Oncol 14:241-247.

Ooe N, Saito K, Mikami N, Nakatuka I, Kaneko H. 2004. Identification of a novel basic helix-loop-helix-PAS factor, NXF, reveals a Sim2 competitive, positive regulatory role in dendriticcytoskeleton modulator drebrin gene expression. Mol Cell Biol 24: 608–616.

Owczarek CM, Portbury KJ, Hardy MP, O'Leary DA, Kudoh J, Shibuya K, Shimizu N, Kola I, Hertzog PJ. 2004. Detailed mapping of the ERG-ETS2 interval of human chromosome 21 and comparison with the region of conserved synteny on mouse chromosome 16. Gene 324:65–77.

Paolini R, Bonaldi L, Bianchini E, Ramazzina E, Cella G. 2003. Spontaneous evolution of essential thrombocythaemia into acute megakaryoblastic leukaemia with trisomy 8, trisomy 21 and cutaneous involvement. Eur J Haematol 71:466-469.

Parker SE, Mai CT, Canfield MA, Rickard R, Wang Y, Meyer RE, Anderson P, Mason CA, Collins JS, Kirby RS, Correa A. 2010. National Birth Defects Prevention Network. Updated National Birth Prevalence estimates for selected birth defects in the United States, 2004–2006. Birth Defects Res A Clin Mol Teratol 88:1008–1016.

Patja K, Pukkala E, Sund R, Iivanainen M, Kaski M. 2006. Cancer incidence of persons with Down syndrome in Finland: a population-based study. Int J Cancer 118:1769-1772.

Patterson D. 2009. Molecular genetic analysis of Down syndrome. Hum Genet 126:195-214.

Pelleri MC, Piovesan A, Caracausi M, Berardi AC, Vitale L, Strippoli P. 2014. Integrated differential transcriptome maps of Acute Megakaryoblastic Leukemia (AMKL) in children with or without Down Syndrome (DS). BMC Med Genomics 7:63.

Pelleri MC, Cicchini E, Locatelli C, Vitale L, Caracausi M, Piovesan A, Rocca A, Poletti G, Seri M, Strippoli P, Cocchi G. 2016. Highly restricted Down syndrome critical region identified on human chromosome 21. Manuscript submitted.

Pennington BF, Moon J, Edgin J, Stedron J, Nadel L. 2003. The neuropsychology of Down syndrome: evidence for hippocampal dysfunction. Child Development 74:75–93.

Piovesan A, Vitale L, Pelleri MC, Strippoli P. 2013. Universal tight correlation of codon bias and pool of RNA codons (codonome): The genome is optimized to allow any distribution of gene expression values in the transcriptome from bacteria to humans. Genomics 101:282–289.

Posch MG, Waldmuller S, Muller M, Scheffold T, Fournier D, Andrade-Navarro MA, De Geeter B, Guillaumont S, Dauphin C, Yousseff D, et al. 2011. Cardiac alpha-myosin (MYH6) is the predominant sarcomeric disease gene for familial atrial septal defects. PLoS One 6:e28872.

Prandini P, Deutsch S, Lyle R, Gagnebin M, Delucinge Vivier C, Delorenzi M, Gehrig C, Descombes P, Sherman S, Dagna Bricarelli F, et al. 2007. Natural gene-expression variation in Down syndrome modulates the outcome of gene-dosage imbalance. Am J Hum Genet 81:252–263.

Pritchard M, Reeves RH, Dierssen M, Patterson D, Gardiner KJ. 2008. Down syndrome and the genes of human chromosome 21: current knowledge and future potentials. Report on the Expert workshop on the biology of chromosome 21 genes: towards gene-phenotype correlations in Down syndrome. Cytogenet Genome Res 121:67-77.

Radtke I, Mullighan CG, Ishii M, Su X, Cheng J, Ma J, Ganti R, Cai Z, Goorha S, Pounds SB, et al. 2009. Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia. Proc Natl Acad Sci U S A 106:12944-12949.

Rahmani Z, Blouin JL, Créau-Goldberg N, Watkins PC, Mattei JF, Poissonnier M, Prieur M, Chettouh Z, Nicole A, Aurias A, et al. 1990. Down syndrome critical region around D21S55 on proximal 21q22.3. Am J Med Genet Suppl 7:98-103

Ramachandran D, Zeng Z, Locke AE, Mulle JG, Bean LJ, Rosser TC, Dooley KJ, Cua CL, Capone GT, Reeves RH, et al. 2015. Genome-Wide Association Study of Down Syndrome-Associated Atrioventricular Septal Defects. G3 (Bethesda) 5:1961-1971.

Raoul O, Carpentier S, Dutrillaux B, Mallet R, Lejeune J. 1976. Partial trisomy of chromosome 21 by maternal translocation t(15;21) (q26.2; q21) [Article in French]. Ann Genet 19:187-190.

Rastelli GC, Kirklin JW, Titus JL. 1966. Anatomic observations on complete form of persistent common atrioventricular canal with special reference to atrioventricular valves. Mayo Clin Proc 41:296 –308.

Rastelli GC, Ongley PA, Kirklin JW, McGoon DC. 1968. Surgical repair of the complete form of persistent common atrioventricular canal. J Thorac Cardiovasc Surg 55:299 –308.

Repeat Masker (http://www.repeatmasker.org/).

Reynolds LE, Watson AR, Baker M, Jones TA, D'Amico G, Robinson SD, Joffre C, Garrido-Urbani S, Rodriguez-Manzaneque JC, Martino-Echarri E, et al., 2010. Tumour angiogenesis is reduced in the Tc1 mouse model of Down's syndrome. Nature 465:813–817.

Ribeiro Ede A Jr., Pinotsis N, Ghisleni A, Salmazo A, Konarev PV, Kostan J, Sjoblom B, Schreiner C, Polyansky AA, Gkougkoulia EA, et al. 2014. The structure and regulation of human muscle alpha-actinin. Cell 159:1447-1460.

Roberts I, Izraeli S. 2014. Haematopoietic development and leukaemia in Down syndrome. Br J Haematol 167:587–599.

Roela RA, Carraro DM, Brentani HP, Kaiano JH, Simao DF, Guarnieiro R, Lopes LF, Borojevic R, Brentani MM. 2007. Gene stage-specific expression in the microenvironment of pediatric myelodysplastic syndromes. Leuk Res 31:579–589.

Roizen NJ, Patterson D. Down's syndrome. 2003. Lancet 361:1281-1289.

Ronn T, Poulsen P, Tuomi T, Isomaa B, Groop L, Vaag A, Ling C. 2009. Genetic variation in ATP5O is associated with skeletal muscle ATP50 mRNA expression and glucose uptake in young twins. PLoS One 4:e4793.

Roper RJ, Reeves RH. 2006. Understanding the basis for Down syndrome phenotypes. PLoS Genet 2:e50.

Roth RB, Hevezi P, Lee J, Willhite D, Lechner SM, Foster AC, Zlotnik A. 2006. Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. Neurogenetics 7:67-80.

Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, Nativ N, Bahir I, Doniger T, Krug H, et al. 2010. GeneCards Version 3: the human gene integrator. Database (Oxford) 2010:baq020.

Sanchez D, Figarella C,Marchand-Pinatel S, Bruneau N, Guy-Crotte O. 2001. Preferential expression of reg I beta gene in human adult pancreas. Biochem Biophys Res Commun 284:729–737.

Sarantitis I, Papanastasopoulos P, Manousi M, Baikoussis NG, Apostolakis E. 2012. The cytoskeleton of the cardiac muscle cell. Hellenic J Cardiol 53:367-379.

Sarkans U, Parkinson H, Lara GG, Oezcimen A, Sharma A, Abeygunawardena N, Contrino S, Holloway E, Rocca-Serra P, Mukherjee G, et al. 2005. The ArrayExpress gene expression database: a software engineering and implementation perspective. Bioinformatics 21:1495–1501.

Sato T, Morishima Y, Shirasaki Y. 2003. 3-[2-[4-(3-Chloro-2-methylphenyl) 21-piperazinyl]ethyl]-5,6-dimethoxy-1-(4-imidazolylmethyl)21H-indazole dihydrochloride 3.5 hydrate (DY-9760e), a novel calmodulin antagonist, reduces brain edema through the inhibition of enhanced blood-brain barrier permeability after transient focal ischemia. J Pharmacol Exp Ther 304:1042–1047.

Sharrocks A. 1994. The design of primer for PCR. In: Griffin HG, Griffin AM, ed. PCR Technology—Current Innovations. Boca Raton, FL: CRC Press, pp 5-11.

Shen YC, Tsai HM, Cheng MC, Hsu SH, Chen SF, Chen CH. 2012. Genetic and functional analysis of the gene encoding neurogranin in schizophrenia. Schizophr Res 137:7–13.

Shirasaki Y, Kanazawa Y, Morishima Y, Makino M. 2006. Involvement of calmodulin in neuronal cell death. Brain Res 1083:189–195.

Sinet PM, Allard D, Lejeune J, Jerome H. Letter. 1975. Gene dosage effect in trisomy 21. Lancet 1:276.

Slungaard A. 2005. Platelet factor 4: a chemokine enigma. Int J Biochem Cell Biol 37:1162–1167.

Solaro RJ, Henze M, Kobayashi T. 2013. Integration of troponin I phosphorylation with cardiac regulatory networks. Circ Res 112:355-366.

Song BG, Jeon ES, Kim YH, Kang MK, Doh JH, Kim PH, Ahn SJ, Oh HL, Kim HJ, Sung JD, et al. 2005. Correlation between levels of N-terminal pro-B-type natriuretic peptide and degrees of heart failure. Korean J Intern Med 20:26-32.

Song P, Rekow SS, Singleton CA, Sekhon HS, Dissen GA, Zhou M, Campling B, Lindstrom J, Spindel ER. 2013. Choline transporterlike protein 4 (CTL4) links to non-neuronal acetylcholine synthesis. J Neurochem 126:451–461.

Sorianello E, Soriano FX, Fernandez-Pascual S, Sancho A, Naon D, Vila-Caballer M, Gonzalez-Navarro H, Portugal J, Andres V, Palacin M, Zorzano A. 2012. The promoter activity of human Mfn2 depends on Sp1 in vascular smooth muscle cells. Cardiovasc Res 94:38-47.

Staedtler F, Hartmann N, Letzkus M, Bongiovanni S, Scherer A, Marc P, Johnson KJ, Schumacher MM. 2013. Robust and tissueindependent gender-specific transcript biomarkers. Biomarkers 18:436–445.

Strippoli P, Pelleri MC, Caracausi M, Vitale L, Piovesan A, Locatelli C, Mimmi MC, Berardi AC, Ricotta D, Radeghieri A et al. 2013. An integrated route to identifying new pathogenesis-based therapeutic approaches for trisomy 21 (Down Syndrome) following the thought of Jérôme Lejeune. Science Postprint 1: e00010.

Sun X, Shin C, Windebank AJ. 1997. Calmodulin in ischemic neurotoxicity of rat hippocampus in vitro. NeuroReport 8:415–418.

Takamatsu K, Maekawa K, Togashi T, Choi DK, Suzuki Y, Taylor TD, Toyoda A, Sugano S, Fujiyama A, Hattori M, Sakaki Y, Takeda T. 2002. Identification of two novel primate-specific genes in DSCR. DNA Res 9:89-97.

Taniguchi Y, London R, Schinkmann K, Jiang S, Avraham H. 1999. The receptor protein tyrosine phosphatase, PTP-RO, is upregulated during megakaryocyte differentiation and is associated with the c-Kit receptor. Blood 94:539–549.

Taub JW, Huang X, Matherly LH, Stout ML, Buck SA, Massey GV, Becton DL, Chang MN, Weinstein HJ, Ravindranath Y. 1999. Expression of chromosome 21-localized genes in acute myeloid leukemia: differences between Down syndrome and non-Down syndrome blast cells and relationship to in vitro sensitivity to cytosine arabinoside and daunorubicin. Blood 94:1393–1400.

Tenedini E, Fagioli ME, Vianelli N, Tazzari PL, Ricci F, Tagliafico E, Ricci P, Gugliotta L, Martinelli G, Tura S, et al. 2004. Gene expression profiling of normal and malignant CD34-derived megakaryocytic cells. Blood 104:3126–3135.

Tomasson MH, Xiang Z, Walgren R, Zhao Y, Kasai Y, Miner T, Ries RE, Lubman O, Fremont DH, McLellan MD, et al. 2008. Somatic mutations and germline sequence variants in the expressed tyrosine kinase genes of patients with de novo acute myeloid leukemia. Blood 111:4797–4808.

Tonelli R, Scardovi AL, Pession A, Strippoli P, Bonsi L, Vitale L, Prete A, Locatelli F, Bagnara GP, Paolucci G. 2000. Compound heterozygosity for two different amino-acid substitution mutations in the thrombopoietin receptor (c-mpl gene) in congenital amegakaryocytic thrombocytopenia (CAMT). Hum Genet 107:225–233.

Trabzuni D, Ramasamy A, Imran S, Walker R, Smith C, Weale ME, Hardy J, Ryten M, North American Brain Expression Consortium. 2013. Widespread sex differences in gene expression and splicing in the adult human brain. Nat Commun 4:2771.

Tu Z, Wang L, Xu M, Zhou X, Chen T, Sun F. 2006. Further understanding human disease genes by comparing with housekeeping genes and other genes. BMC Genomics 7:31.

Tunnacliffe A, Majumdar S, Yan B, Poncz M. 1992. Genes for beta-thromboglobulin and platelet factor 4 are closely linked and form part of a cluster of related genes on chromosome 4. Blood 79:2896–2900.

Tunstall-Pedoe O, Roy A, Karadimitris A, de la Fuente J, Fisk NM, Bennett P, Norton A, Vyas P, Roberts I. 2008. Abnormalities in the myeloid progenitor compartment in Down syndrome fetal liver precede acquisition of GATA1 mutations. Blood 112:4507-4511.

University of California Santa Cruz (UCSC) Genome Browser. (http://genome-euro.ucsc.edu/cgi-bin/hgGateway).

Vaughan CJ, Basson CT. 2000. Molecular determinants of atrial and ventricular septal defects and patent ductus arteriosus. Am J Med Genet 97:304-9.

Vawter MP, Evans S, Choudary P, Tomita H, Meador-Woodruff J, Molnar M, Li J, Lopez JF, Myers R, Cox D, et al. 2004. Gender-specific gene expression in postmortem human brain: localization to sex chromosomes. Neuropsychopharmacology 29:373–384.

Vitale L, Casadei R, Canaider S, Lenzi L, Strippoli P, D'Addabbo P, Giannone S, Carinci P, Zannotti M. 2002. Cysteine and tyrosine-rich 1 (CYYR1), a novel unpredicted gene on human chromosome 21 (21q21.2), encodes a cysteine and tyrosine-rich protein and defines a new family of highly conserved vertebrate-specific genes. Gene 290:141-151.

Vitale L, Frabetti F, Huntsman SA, Canaider S, Casadei R, Lenzi L, Facchin F, Carinci P, Zannotti M, Coppola D, Strippoli P. 2007. Sequence, "subtle" alternative splicing and expression of the CYYR1 (cysteine/tyrosine-rich 1) mRNA in human neuroendocrine tumors. BMC Cancer 7:66.

Wang C, Gong B, Bushel PR, Thierry-Mieg J, Thierry-Mieg D, Xu J, Fang H, Hong H, Shen J, Su Z et al. 2014. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nat Biotechnol 32:926-932.

Wang H, Iacoangeli A, Popp S,Muslimov IA, Imataka H, Sonenberg N, Lomakin IB, Tiedge H. 2002. Dendritic BC1 RNA: functional role in regulation of translation initiation. J Neurosci 22:10232–10241.

Wang PJ, McCarrey JR, Yang F, Page DC. 2001. An abundance of X-linked genes expressed in spermatogonia. Nat Genet 27:422–426.

Wang X, Zhao Y, Zhang X, Badie H, Zhou Y, Mu Y, Loo LS, Cai L, Thompson RC, Yang B et al. 2013. Loss of sorting nexin 27 contributes to excitatory synaptic dysfunction by modulating glutamate receptor recycling in Down's syndrome. Nat Med 19:473-480.

Wechsler J, Greene M, McDevitt MA, Anastasi J, Karp JE, Le Beau MM, Crispino JD. 2002. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. Nat Genet 32:148-152.

Williamson R, van Aalten L, Mann DM, Platt B, Plattner F, Bedford L, Mayer J, Howlett D, Usardi A, Sutherland C, Cole AR. 2011. CRMP2 hyperphosphorylation is characteristic of Alzheimer's disease and not a feature common to other neurodegenerative diseases. J Alzheimers Dis 27:615–625.

Xu AG, He L, Li Z, Xu Y, Li M, Fu X, Yan Z, Yuan Y, Menzel C, Li N, et al. 2010. Intergenic and repeat transcription in human, chimpanzee and macaque brains measured by RNA-Seq. PLoS Comput Biol 6: e1000843.

Xu ZP, Dutra A, Stellrecht CM, Wu C, Piatigorsky J, Saunders GF. 2002. Functional and structural characterization of the human gene BHLHB5, encoding a basic helix-loop-helix transcription factor. Genomics 80:311–318.

Yagi T, Morimoto A, Eguchi M, Hibi S, Sako M, Ishii E, Mizutani S, Imashuku S, Ohki M, Ichikawa H. 2003. Identification of a gene expression signature associated with pediatric AML prognosis. Blood 102:1849-1856.

Yokoyama M, Nishi Y, Yoshii J, Okubo K, Matsubara K. 1996. Identification and cloning of neuroblastoma-specific and nerve tissue-specific genes through compiled expression profiles. DNA Res 3:311–320.

Yokoyama T, Toki T, Aoki Y, Kanezaki R, Park MJ, Kanno Y, Takahara T, Yamazaki Y, Ito E, Hayashi Y, Nakamura T. 2012. Identification of TRIB1 R107L gain-of-function mutation in human acute megakaryocytic leukemia. Blood 119:2608–2611.

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, et al. 2002. A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296:79-92.

Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. 2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One 9:e78644.

Zhou L, Zhang R,Wang L, Shen S, Okamoto H, Sugawara A, Xia L, Wang X, Noguchi N, Yoshikawa T, et al. 2010. Upregulation of REG Ialpha accelerates tumor progression in pancreatic cancer with diabetes. Int J Cancer 127:1795–1803.

Zhu H, Chen L, Zhou W, Huang Z, Hu J, Dai S, Wang X, Huang X, He C. 2013. Over-expression of the ATP5J gene correlates with cell migration and 5-fluorouracil sensitivity in colorectal cancer. PLoS One 8:e76846.

Zunszain PA, Anacker C, Cattaneo A, Choudhury S, Musaelyan K, Myint AM, Thuret S, Price J, Pariante CM. 2012. Interleukin-1β: a new regulator of the kynurenine pathway affecting human hippocampal neurogenesis. Neuropsychopharmacology 37:939–949.

Pubblications

Strippoli P, Pelleri MC, Caracausi M, Vitale L, Piovesan A, Locatelli C, Mimmi MC, Berardi AC, Ricotta D, Radeghieri A, Barisani B, Basik M, Monaco MC, Ghezzo A, Marco Seri M and Guido Cocchi G. 2013. An integrated route to identifying new pathogenesis-based therapeutic approaches for trisomy 21 (Down Syndrome) following the thought of Jérôme Lejeune. Science Postprint 1(1): e00010.

Piovesan A, Caracausi M, Pelleri MC, Vitale L, Martini S, Bassani C, Gurioli A, Casadei R, Soldà G, Strippoli P. 2014. Improving mRNA 5' coding sequence determination in the mouse genome. Mamm Genome 25:149-159.

Caracausi M, Vitale L, Pelleri MC, Piovesan A, Bruno S, Strippoli P. 2014. A quantitative transcriptome reference map of the normal human brain. Neurogenetics 15:267-287.

Pelleri MC, Piovesan A, Caracausi M, Berardi AC, Vitale L, Strippoli P. 2014. Integrated differential transcriptome maps of Acute Megakaryoblastic Leukemia (AMKL) in children with or without Down Syndrome (DS). BMC Med Genomics 7:63.

Piovesan A, Caracausi M, Ricci M, Strippoli P, Vitale L, Pelleri MC. 2015. Identification of minimal eukaryotic introns through GeneBase, a user-friendly tool for parsing the NCBI Gene databank. DNA Res. 22:495-503.

Caracausi M, Rigon V, Piovesan A, Strippoli P, Vitale L, Pelleri MC. 2016. A quantitative transcriptome reference map of the normal human hippocampus. Hippocampus. 26:13-26.

Pelleri MC, Cicchini E, Locatelli C, Vitale L, Caracausi M, Piovesan A, Rocca A, Poletti G, Seri M, Strippoli P, Cocchi G. 2016. Highly restricted Down syndrome critical region identified on human chromosome 21. Manuscript submitted.

Caracausi M, Piovesan A, Vitale L, Pelleri MC. 2016. Integrated transcriptome map highlights structural and functional aspects of the normal human heart. Manuscript submitted.