

Research article

Open Access

## Detection of chromosomal regions showing differential gene expression in human skeletal muscle and in alveolar rhabdomyosarcoma

Andrea Bisognin, Stefania Bortoluzzi and Gian Antonio Danieli\*

Address: Department of Biology, University of Padua, via Ugo Bassi 58B, 35131, Padova, Italy

Email: Andrea Bisognin - bisio@medgen.univr.it; Stefania Bortoluzzi - stefibo@bio.unipd.it; Gian Antonio Danieli\* - danieli@bio.unipd.it

\* Corresponding author

Published: 03 June 2004

Received: 19 December 2003

BMC Bioinformatics 2004, 5:68 doi:10.1186/1471-2105-5-68

Accepted: 03 June 2004

This article is available from: <http://www.biomedcentral.com/1471-2105/5/68>

© 2004 Bisognin et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** Rhabdomyosarcoma is a relatively common tumour of the soft tissue, probably due to regulatory disruption of growth and differentiation of skeletal muscle stem cells. Identification of genes differentially expressed in normal skeletal muscle and in rhabdomyosarcoma may help in understanding mechanisms of tumour development, in discovering diagnostic and prognostic markers and in identifying novel targets for drug therapy.

**Results:** A Perl-code web client was developed to automatically obtain genome map positions of large sets of genes. The software, based on automatic search on Human Genome Browser by sequence alignment, only requires availability of a single transcribed sequence for each gene. In this way, we obtained tissue-specific chromosomal maps of genes expressed in rhabdomyosarcoma or skeletal muscle. Subsequently, Perl software was developed to calculate gene density along chromosomes, by using a sliding window. Thirty-three chromosomal regions harbouring genes mostly expressed in rhabdomyosarcoma were identified. Similarly, 48 chromosomal regions were detected including genes possibly related to function of differentiated skeletal muscle, but silenced in rhabdomyosarcoma.

**Conclusion:** In this study we developed a method and the associated software for the comparative analysis of genomic expression in tissues and we identified chromosomal segments showing differential gene expression in human skeletal muscle and in alveolar rhabdomyosarcoma, appearing as candidate regions for harbouring genes involved in origin of alveolar rhabdomyosarcoma representing possible targets for drug treatment and/or development of tumor markers.

### Background

Rhabdomyosarcoma (RMS) is the most common sarcoma of the soft-tissue, the third most common extracranial solid tumor in children [1]. The incidence of RMS is about four new cases per million per year [2]. RMS is believed to arise from regulatory disruption of growth and differentiation of skeletal muscle stem cells [3,4].

Several approaches, such as FISH, Comparative Genomic Hybridization (CGH), representational difference analysis [5-7] were proposed to investigate at a genomic level DNA amplification in tumour samples. Microarray technology methods, applied to identification of genome-wide chromosomal imbalances tried to overcome limitations of conventional CGH by hybridizing sample DNA to mapped sequences instead of metaphase chromosomes

[8]. More recently, comparative expressed sequence hybridisation (CESH) was applied to the analysis of chromosomal regions including genes overexpressed in a leukemic cell line and in four rhabdomyosarcoma cell lines [9]. A number of discrete regions of increased tumour expression (RITEs) have been identified by a computational method, which compared gene expression levels in unbiased EST libraries from tumour and normal tissues. Several RITEs were identified, corresponding to chromosomal segments previously detected by CGH analysis. Permutation analyses suggested that chromosomal and genomic distribution of RITEs is probably not random [10]. Unfortunately, this study did not consider muscle tissues or muscle tumours.

The aim of this study was to build and compare transcript maps of normal human skeletal muscle and of alveolar rhabdomyosarcoma, in order to identify specific regions of the human genome involved in development of this tumour. This would facilitate discovery of novel genes playing a role in rhabdomyosarcoma and, possibly, of novel tumour markers.

In this paper we report on detection of chromosomal regions harbouring genes mostly expressed in RMS or downregulated in RMS, if compared with normal skeletal muscle, by using different computational methods.

## Results and Discussion

### **Reconstruction of expression profiles at genome level**

We reconstructed tissue expression profiles by a computational approach, starting from unbiased cDNA libraries data. In particular, the expression profile of normal adult skeletal muscle (SM) was reconstructed by using information from 26,964 ESTs, corresponding to 8,517 genes expressed in SM, such as from data obtained by merging four cDNA libraries. Human alveolar rhabdomyosarcoma (ARMS) expression profile was reconstructed from 24,175 ESTs, derived from one library, accounting for 4,549 genes expressed in ARMS. Full data are accessible online [11].

The number of ESTs reported in each UniGene cluster recorded in the expression profile was used to estimate the level of activity of the corresponding gene, as previously described [12], under the assumption that the number of detected ESTs per gene is a function of its transcript frequency in the original population of messenger RNA. Thus, each expression profile is a catalogue of genes expressed in a given tissue. Each gene in the catalogue is identified by UniGene cluster ID and RefSeq accession number and it is associated to the encoded product description and linked to LocusLink and to GenBank databases.

### **Estimation of density of expressed genes along chromosomes**

Genomic map positions, established at one-nucleotide level of resolution, were obtained by querying UCSC BLAT facility with representative DNA sequences of each of 8,517 genes expressed in skeletal muscle and of 5,520 genes expressed in ARMS, according to RefSeq or to UniGene database. Location of each gene in the human genome sequence is given by starting and ending position of the transcribed region, expressed as bp distances from the short arm telomere.

After stringent alignment quality check (see Methods) 98% of genes considered by the study were unambiguously located to their chromosomal positions. In particular, the precise chromosomal location was determined for 8,308 genes expressed in SM and for 5,414 genes expressed in ARMS.

Chromosomal base-pair coordinates of such genes were used to establish "gene density" along each chromosome. Gene density was expressed as a fraction of the genomic sequence covered by gene-associated sequences, in chromosomal regions spanned by a sliding window. Gene density of skeletal muscle genes and of ARMS genes was calculated using 1 Mb-wide windows with 10 Kb overlap, thus obtaining 304,691 different values of gene density along the whole human genome. Window width was set to 1 Mb, in order to scan the genome with a window wider than the average length of most human genes. Taking into account that the average size of human genes is about 14 Kb and the median value is 27 Kb [13], that gene density is highly variable in the human genome, ranging from zero to about 100 genes per Mb (e.g. 6p21.33), and that only a fraction of all human genes is expressed in a given tissue, the selected dimension of one megabase for the window used in the present study should be adequate for measuring gene density in the human genome. The shift between adjacent windows was set to 10 Kb, in order to obtain sufficiently numerous data points, for a precise calculation of gene density along human chromosomes.

### **Comparison between density of genes expressed in normal skeletal muscle and ARMS**

Gene density values, normalized to the total number of expressed genes, were plotted against the base-pair scale length of different human autosomes and of the X chromosome.

Observed gene density values ranged from zero to 1.993 for SM and from zero to 1.197 for ARMS. Average gene density were 0.129 and 0.069 for SM and ARMS, respectively. The difference between normalised ARMS and normal skeletal muscle gene densities was measured. For 27.6% of the genome ARMS gene density was higher than

**Table 1: Muscle and RMS gene density in the human genome.**

Chr.	Chromosomal regions			Diff. expr. in RMS	Chr.	Chromosomal regions			Diff. expr. in RMS
	Start(Mb)	End(Mb)	Cyt. band			Start(Mb)	End(Mb)	Cyt. band	
<b>1</b>	14.09	14.75	p36.21*	+	<b>13</b>	59.57 94.91	59.69 95.14	q21.2 q32.1	+ -
	50.16	50.79	p32.3	+					
	63.86	64.11	p31.3	+					
	91.99	92.09	p22.1	-					
	96.84	97.50	p21.3	-					
	173.14	173.59	q25.2	-					
<b>2</b>	<b>29.21</b>	<b>30.22</b>	<b>p23.2 – p23.1*</b>	+	<b>14</b>	65.07 71.51	65.58 71.57	q23.3 q24.2	- -
	<b>50.05</b>	<b>51.22</b>	<b>p16.3</b>	-					
	<b>79.57</b>	<b>81.03</b>	<b>p12</b>	-					
	99.66	100.17	q11.2	-					
	137.73	138.53	q22.1	+					
	143.91	144.65	q22.2-q22.3	+					
	<b>205.48</b>	<b>206.28</b>	<b>q33.3</b>	-					
	<b>212.32</b>	<b>213.26</b>	<b>q34</b>	-					
<b>3</b>	<b>2.03</b>	<b>3.16</b>	<b>p26.3-p26.2</b>	+	<b>15</b>	22.97	23.04	q11.2	-
	29.28	29.78	p24.1	-					
	37.17	37.19	p22.3	-					
	37.24	37.61	p22.3	-					
	<b>53.92</b>	<b>55.06</b>	<b>p21.1-p14.3*</b>	+					
	61.51	62.03	p14.2	-					
	76.79	77.61	p12.3*	+					
	101.86	102.32	q12.2-q12.3	-					
	116.71	117.55	q13.31	+					
	160.38	160.91	q25.33	-					
	176.08	176.82	q26.31	-					
<b>4</b>	53.52	54.15	q12	+	<b>16</b>	<b>82.50</b> 88.74	<b>83.51</b> 88.88	<b>q23.3</b> q24.3*	- +
	87.50	88.06	q21.3	-					
	114.54	114.90	q25-q26	-					
	151.75	152.30	q31.23-q31.3	-					
	162.88	163.45	q32.2	-					
	168.14	168.84	q32.3	+					
<b>5</b>	19.37	20.08	p14.3	+	<b>17</b>	44.45 45.08	44.72 45.13	q21.31 q21.31	- -
	89.80	90.59	q14.3	+					
	92.87	93.54	q15	+					
	168.12	168.58	q34-q35.1	-					
<b>6</b>	<b>1.51</b>	<b>2.39</b>	<b>p25.3-p25.2</b>	+	<b>18</b>	7.66 <b>49.86</b>	8.29 <b>50.84</b>	p11.23 <b>q21.1-q21.2</b>	- -
	20.70	21.19	p22.3	-					
	37.94	38.43	p21.2	-					
	44.67	45.37	p21.1*	+					
	69.40	70.07	q12-q13	-					
	128.69	129.74	q22.33	-					
	152.69	153.01	q25.2	+					
<b>7</b>	36.54	37.32	p14.2-p14.1	+	<b>19</b>	38.48	38.54	q13.11*	+
	<b>68.36</b>	<b>69.76</b>	<b>q11.22*</b>	+					
	100.90	101	q22.1*	+					
	109.95	110.64	q31.1	-					
	132.56	133.21	q33	-					
<b>8</b>	98.97	99.41	q22.2	-	<b>20</b>	8.18	8.71	p12.3	-

**Table 1: Muscle and RMS gene density in the human genome. (Continued)**

<b>9</b>	8.41	8.90	p24.1	-	<b>21</b>	16.32	16.61	q21.1	-
	109.93	109.99	q32*	+					
	<b>112.29</b>	<b>113.53</b>	<b>q32-q33.1*</b>	+					
	121.63	122.47	q33.3*	+					
<b>10</b>	0.50	0.75	p15.3*	+	<b>22</b>	31.87	32.74	q12.3*	+
	<b>52.83</b>	<b>53.84</b>	<b>q21.1</b>	-					
	75.66	76.54	q22.2*	+					
	78.64	79.18	q22.3	-					
	94.41	94.48	q23.33	-					
	108.33	109.16	q25.1	+					
<b>11</b>	10.20	10.28	p15.4	-	<b>X</b>	8.74	9.47	p22.31-p22.22*	+
	10.34	10.34	p15.4	-					
	<b>20.60</b>	<b>21.90</b>	<b>q22.3</b>	+					
	<b>83.50</b>	<b>84.83</b>	<b>q14.1</b>	-					
<b>12</b>	26.85	26.92	p11.23	-	<b>Y</b>	-	-	-	-

Local density of genes expressed in skeletal muscle and in RMS was calculated on the basis of 304,691 sliding windows, 1 Mb-wide and with 10 Kb overlap. Chromosomal regions in which the absolute difference between RMS and skeletal muscle gene density was over 0.6 were selected. Thirty-three chromosomal regions harbouring mostly genes expressed in RMS and 48 chromosomal segments, possibly related to function of differentiated skeletal muscle but silent in RMS are reported. Regions including chromosomal segments in which absolute difference between ARMS and skeletal muscle gene density exceeded 1.0 are in bold, whereas asterisks indicate regions in which overexpression of tumour genes has been reported for different tissues by Zhou and colleagues [10].

SM. The absolute value of gene density difference was higher than 0.3 for 10,653 windows (7.6%). A total of 2,384 windows showing over 0.6 absolute difference were observed (0.7%), corresponding to 33 chromosomal regions harbouring genes overexpressed in ARMS if compared with SM and to 48 chromosomal segments, which appeared strongly downregulated in ARMS (Table 1 and Figure 2). The 0.6 threshold for the observed difference of gene density has been selected in order to pick up genomic regions in which the difference was either moderate or high, without producing an unmanageable increase of genomic regions included and, in particular, to avoid their fragmentation.

Complete plots of gene densities along different human chromosomes are available as supplementary material at our website [14]. As an example, details of the human X chromosome are shown in Figure 1.

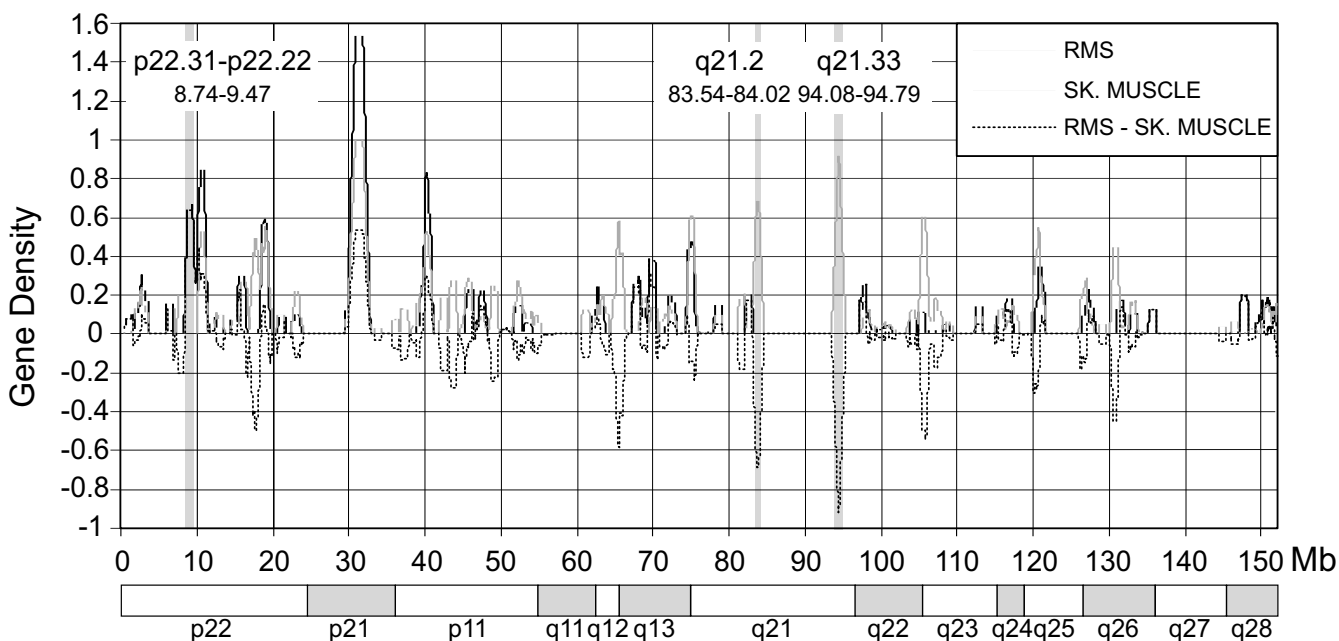
Identification of genes differentially expressed in normal and tumour tissues may provide a valuable help in understanding tumour development, in discovering diagnostic and prognostic markers and in identifying novel targets for drug therapy. Digital expression profiling of six different solid tumour types has been used for discovering novel cancer genes [15], but the data set included both normalized and subtracted cDNA libraries and Fisher's exact test was inappropriately used. Artificial neural networks have been used to classify different cancer samples or cell lines on the basis of their expression signature [16].

As far as ARMS is concerned, a cDNA microarray study showed consistent patterns of gene expression in different ARMS cell lines [17].

In the present study, by associating transcripts to the corresponding gene sequence on the human genome, we mapped at genomic level genes expressed in SM and in ARMS. Moreover, we calculated density along each human chromosome of such genes, by using a sliding window approach. Because of the relatively low gene density per window, a statistical approach for comparing gene density was considered to be less convenient than a "qualitative", plot-based approach.

In human genome, possible clustering of genes highly or specifically expressed in specific tissues has been suggested by different studies (see for instance [18-20]) which considered both normal and cancer tissues. Fujii and colleagues [21] found that numerous groups of genes expressed in specific tumours, such as squamous cell carcinoma or adenocarcinoma, cluster in given regions of the human genome. On the other hand, recurrent non random imbalances pertaining specific chromosomal regions are frequently reported in tumours. Several experimental or computational methods are currently used for identification of such regions.

We identified specific chromosomal regions hosting more genes expressed in ARMS than expressed in SM. This provides a starting point for identifying novel candidate can-



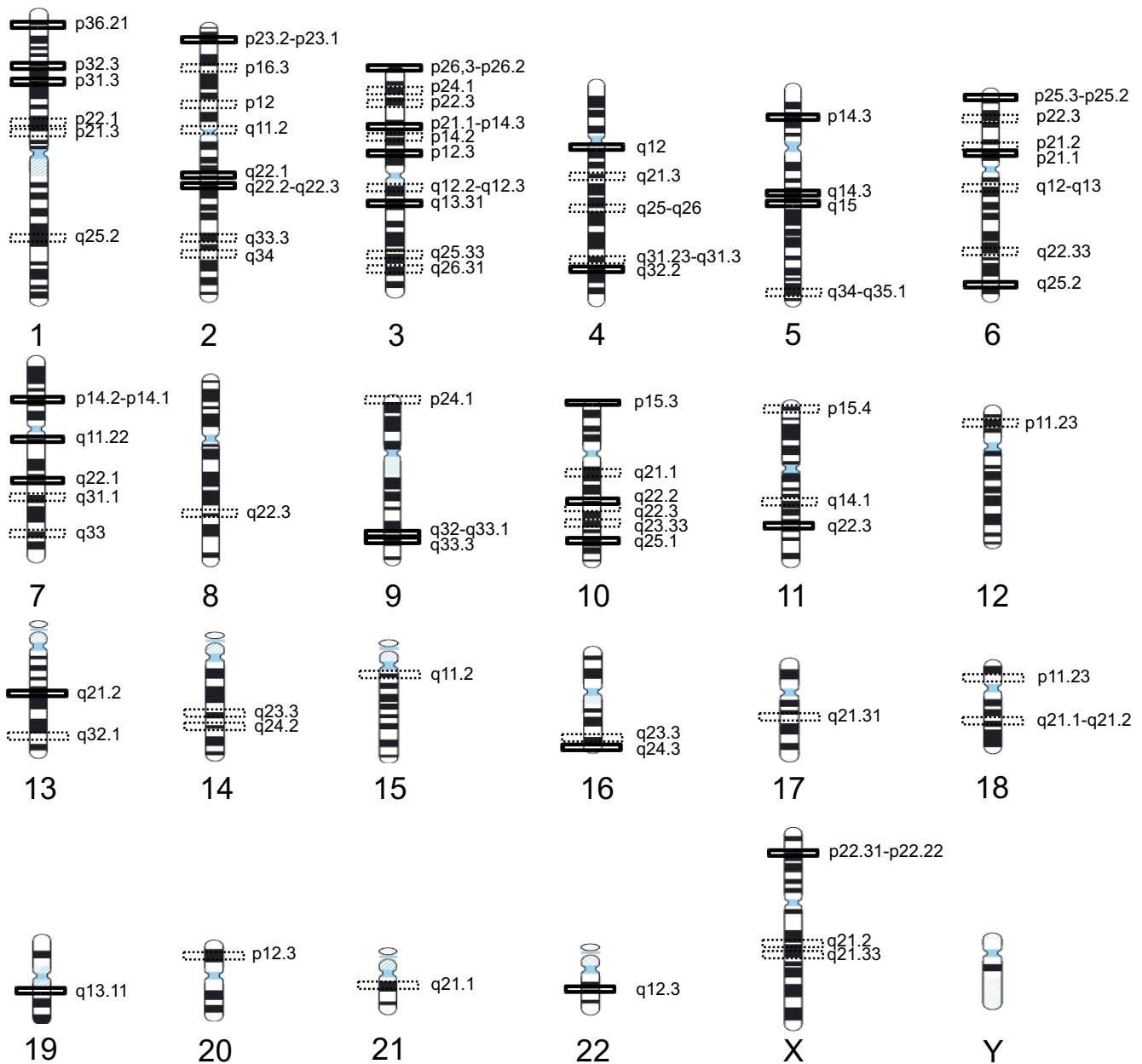
**Figure 1**  
**Density of skeletal muscle and of RMS genes associated sequences along human X chromosome.** Gene density was calculated in chromosomal regions spanned by sliding windows, as the fraction of genomic sequence covered by gene-associated sequences, on the basis of 15,165 windows of 1 Mb with an overlap between adjacent windows of 10 Kb. As an example, details about gene density on the human X chromosome are shown. Gene density values were normalized to the total number of expressed genes and plotted together on the same base-pair scale axis. Chromosomal regions in which absolute difference between RMS and skeletal muscle gene density was over 0.6 were selected. On the X chromosome, one region resulted to harbour mostly genes expressed in RMS (p22.31-p22.22) and two regions resulted to contain genes expressed in fully differentiated muscle but silent in RMS (q21.2 and q21.33). Complete plots of gene densities along all human chromosomes are available as supplementary material [14].

cer genes. On the other hand, comparison of chromosomal maps of genes expressed respectively in SM or in ARMS revealed the existence of genome regions possibly involved in muscle differentiation. Further analysis of these regions could help in understanding why ARMS cells fail to differentiate.

In regions harboring more genes expressed in ARMS than in normal muscle a number of novel genes are present, such as genes with uncharacterized products, hypothetical proteins or transcribed sequences. Among known genes, some are particularly interesting, such as genes encoding for transcription factors or for proteins involved in cell cycle control, in signal transduction or in cell adhesion. For some of them, an involvement in tumors has been already reported. Nine genes encoding transcriptional factors and/or regulators of cell cycle lie in chromosomal regions harboring mostly genes expressed in RMS (DMRT-like family A2, 1p32.3; Transcription factor SMIF, 3p21.1-p14.3; Zinc finger protein 288, 3q13.31; Suppressor of Ty

3 homolog, 6p21.1; Zinc finger, HIT domain containing 1, 7q22.1; Tripartite motif-containing 32, 9q32.-q33.1; Pre-B-cell leukemia transcription factor 3, 9q33.3; PPP3CB: Protein phosphatase 3, catalytic subunit, beta isoform, 10q22.2; Adenovirus 5 E1A binding protein, 10p15.3 and NEL-like 1, 11q22.3). Four genes encoding for adhesion molecules, are also located in regions overexpressed in RMS (Contactin 4; Cadherin 18, Monogenic, audiogenic seizure susceptibility 1 homolog and Vinculin, found in 3p26.3-p26.2; 5p14.3, 5q14.3 and 10q22.2, respectively). Moreover, two oncogenes (Anaplastic lymphoma kinase and Nucleophosmin) map in 2p23.2-p23.1, whereas the "upregulated in colorectal cancer gene 1" maps in 7p14.2-p14.1 and the gene for the inhibitor of metalloproteinase 3, whose expression is induced in response to mitogenic stimulation, maps in 22q12.3.

An effort was made to compare our results with CESH data [9], but the two approaches are different in terms of



**Figure 2**  
**Regions of human chromosomes with absolute difference between RMS and skeletal muscle gene density over 0.6.** Thirty-three chromosomal regions harboring mostly genes expressed in RMS are boxed with continue line, whereas 48 chromosomal segments, possibly related to function of differentiated skeletal muscle but silent in RMS are indicated by dashed boxes.

sensitivity. At its highest definition CESH measures hybridization intensities on a 450-bands resolution metaphase, whereas our procedure localizes genes on the human genome sequence, at single nucleotide level.

However, it may be noticed that in the present study we identified with high confidence one chromosomal region (2p23.2-p23.1) containing a segment showing considerable differential gene density between RMS and SM, within a chromosomal region in which overexpression of RMS

genes was detected by CESH [9]. In addition, in the long arm of chromosome 6, reportedly underexpressed in all ARMS lines analyzed by CESH, we found four different regions harboring genes expressed in SM but silenced in ARMS.

From a technical point of view, comparison with the work reporting detection of RITES [10] was more feasible and meaningful even if they did not take in consideration any SM or ARMS library. We compared chromosomal regions harboring genes more expressed in ARMS than in SM with the list of regions showing increased gene expression in tumour of brain, breast, liver, and lung, identified by Zhou and colleagues [10]. We found that many regions showing higher gene densities in ARMS than in SM were included in this list of regions of increased expression in tumours of several tissues (these regions are indicated with asterisks in Table 1). Therefore, some of these regions could be associated to a general neoplastic phenotype, consequent to genomic rearrangements or due to deregulation of expression of adjacent genes.

Chromosomal regions where differential expression of genes was noticed in ARMS in this study, but were never reported previously, appear as candidate regions for harboring genes possibly involved in origin of ARMS or possible targets for chemotherapy.

Several recurring cytogenetical abnormalities have been observed by Comparative Genomic Hybridization (CGH) studies in RMS cell lines. They involved the gain or loss of complete chromosomes or of specific chromosomal regions. In spite of the different nature of these data, we attempted to find possible overlap between the location of chromosomal segments apparently rich or poor in genes expressed in RMS, identified by the present study, with those of chromosomal regions reportedly involved in prominent imbalances observed in rhabdomyosarcoma by several CGH studies. Several correspondences between ESTs based and CGH results were observed, such as overexpression of RMS genes in chromosomal regions reportedly amplified or translocated in RMS, and underexpression of RMS genes in genomic regions deleted in RMS. In particular, overexpression of genes in ARMS, in regions 1p31.3, 3p12.3, 22q12.3 and 7p14.2-14.1, is consistent respectively with 1p31-p21, 3p12 and 22q amplification and gain of 7p reported by Gordon and colleagues [22]. Underexpression of ARMS genes in regions, 3p24.1, 3p22.3, 3p14.2, 5q34-q32.1, 10q23.33 and 13q32.1 is consistent with loss of 3p, 5q32-qter, 10q23 and chromosome 13 [22]. Furthermore, gain of 2q, reported in 40% of 45 different samples of alveolar and embryonal RMS, 7q (31%), 11q (31%) and 16q (27%) [23] is consistent with our finding of overexpression in RMS of regions 2q22.1 and 2q22.2-q22.3, 7q11.22 and

7q22.1, 11q22.3 and 16q23.3. Moreover, Pandita and colleagues [24] reported gain of regions of chromosomes 2, 5, 7, 8, 11 and 12 in the majority of considered RMS, which is consistent with our findings of eleven distinct regions in these chromosomes harbouring genes expressed in ARMS but silent in SM.

## Conclusions

In this study we developed a method and the associated software for the comparative analysis of gene expression in genome. All developed software is freely available online [14]. We identified chromosomal segments showing differential gene expression in human skeletal muscle and in alveolar rhabdomyosarcoma, which appear as candidate regions for harboring genes involved in origin of ARMS and for being possible targets for drug treatment or markers of tumour development.

## Methods

We considered for the study only "unbiased" UniGene cDNA libraries pertaining to selected human tissues, such as cDNA clones collections which did not undergo, during their construction, normalization or subtraction processes. These methods deeply bias ESTs frequencies and alter correlation between frequency of ESTs pertaining to a specific gene in the library and gene expression level in the considered tissue.

The expression profile of normal adult skeletal muscle (SM) was reconstructed by pooling four cDNA libraries (LibIDs: 45, 530, 6761, 9692). Human alveolar rhabdomyosarcoma (ARMS) expression profile was reconstructed from one very large cDNA library (LibID: 3714).

All data presented in this paper were mined from UniGene release #159 [25], by perl software designed for completing a fully automated procedure for data retrieval and expression profiles reconstruction. After downloading human UniGene clusters and libraries data, expression profiles are reconstructed by eventual pooling of libraries and calculation of expression data. HTML pages are automatically built. Each gene in a profile is identified by gene name and description, UniGene cluster, LocusLink number and GenBank ID of the longest sequence representative of the cluster.

The number of ESTs per each UniGene cluster, pertaining to a specific gene in a given library (or pool of libraries), was used to estimate the gene expression level as a percentage of the total detected transcriptional activity in the tissue (total number of ESTs per library or pool of libraries)[11]. The whole computational work was accomplished by a software pipeline estimating the level of expression of genes and producing expression profiles, integrating gene expression and annotation data in HTML

format with links to external resources as LocusLink and GenBank databases. The interactive console tool is conceived to give the possibility of automatically download of large UniGene files and of lists of cDNA libraries files. By simple commands expression profiles are produced. Different profiles could also be merged to create expression data matrices.

A perl-code web client was developed to automatically obtain genome map positions of large sets of genes, requiring only the availability of a transcribed sequence for each gene. Gene location was established at a nucleotide level resolution, by iteratively querying BLAT search facility at UCSC [26] with the representative nucleotide sequence. Representative sequences of UniGene clusters, or, whenever available, reference sequences of corresponding genes, as established by NCBI RefSeq project [27], were used. BLAT, after searching for similarity between the query sequence and the human genome assembly, outputs the best scoring alignments. We defined stringent thresholds for extension and identity of alignments reported between genomic and query nucleotide sequences, in order to exclude spurious results: only alignments extending for more than 400 bp with an identity percentage greater than 95% were considered. In this way, we obtained tissue-specific chromosomal maps of genes expressed in rhabdomyosarcoma or muscle, encompassing the whole genome. The web client takes as input file a list of UniGene clusters and needs the Hs.data and Hs.sequniq files and their indexes. For each gene, gene name and sequence could be retrieved and this information is used to BLAT the sequence to the Human Genome Sequences and to produce a file output with the list of UniGene clusters genomic positions.

Perl software was also developed to calculate gene density along chromosomes, by using a sliding window. Basically, using as input a list of chromosome lengths in bp and a list of gene locations (start and end of transcribed regions, as bp distances from the p telomere) and given selected window size and length of overlap between adjacent windows, the software allow the calculation of gene density along chromosomes with the method of sliding windows. The calculated gene density for a specific window is associated to its central point and, ultimately, a list of gene densities is compiled.

For each of reconstructed expression profiles, a list of sequences, corresponding to the set of detected transcripts, and the base-pair coordinates corresponding to their start and ending points on the human genome assembly, were used to calculate, for each profile, the fraction of the genome covered by gene associated sequences. A density value, calculated on the selected window length, was therefore associated to the point coordinate of the

human genome sequence corresponding to the center of that window. In this study, gene densities were calculated by using a window length of 1 Mb and a shift of 10 Kb. In presence of overlapping genes, the contribution of each one to the coverage was considered separately. Therefore, the calculated "gene density" resulted to be more than one in some regions.

### Authors' contributions

AB and SB participated in the design of the study and developed the software. GAD participated in the design and coordination of the study. All authors have read and approved this manuscript.

### Acknowledgements

The financial support of the Ministry for the Technological and Scientific Research and of the Padova University to professor GAD is gratefully acknowledged, together with the funding of the Italian Association for Cancer Research (AIRC) to SB. AB is recipient of an AIRC fellowship and SB is recipient of a Post Doc fellowship of the Padova University.

### References

1. Dagher R, Helman L: **Rhabdomyosarcoma: an overview.** *Oncologist* 1999, **4**:34-44.
2. Ruymann FB: **Rhabdomyosarcoma in children and adolescents. A review.** *Hematol Oncol Clin North Am* 1987, **1**:621-54.
3. Merlino G, Helman LJ: **Rhabdomyosarcoma – working out the pathways.** *Oncogene* 1999, **18**:5340-8.
4. Astolfi A, De Giovanni C, Landuzzi L, Nicoletti G, Ricci C, Croci S, Scopece L, Nanni P, Lollini PL: **Identification of new genes related to the myogenic differentiation arrest of human rhabdomyosarcoma cells.** *Gene* 2001, **274**:139-49.
5. Pinkel D, Straume T, Gray JW: **Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization.** *Proc Natl Acad Sci U S A* 1986, **83**:2934-8.
6. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D: **Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors.** *Science* 1992, **258**:818-21.
7. Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD: **Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries.** *Proc Natl Acad Sci U S A* 1996, **93**:6025-30.
8. Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG: **High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.** *Nat Genet* 1998, **20**:207-11.
9. Lu YJ, Williamson D, Clark J, Wang R, Tiffin N, Skelton L, Gordon T, Willis SM R, Allan B, Jackman A, Cooper C, Pritchard-Jones K, Shipley J: **Comparative expressed sequence hybridization to chromosomes for tumour classification and identification of genomic regions of differential gene expression.** *Proc Natl Acad Sci USA* 2001, **98**:9197-9202.
10. Zhou Y, Luoh SM, Zhang Y, Watanabe C, Wu TD, Ostland M, Wood WI, Zhang Z: **Genome-wide identification of chromosomal regions of increased tumor expression by transcriptome analysis.** *Cancer Res* 2003, **63**:5781-5784.
11. **HGXP (Human Gene Expression Profiles)** [<http://telethon.bio.unipd.it/bioinfo/HGXP/>]
12. Bortoluzzi S, d'Alessi F, Danielli GA: **A computational reconstruction of the adult human heart transcriptional profile.** *J Mol Cell Cardiol* 2000, **32**:1931-8.
13. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-



- Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, RSMer J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordtsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korfi I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, Szustakowski J, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, International Human Genome Sequencing Consortium: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
14. **Supplementary material** [<http://humgen.bio.unipd.it/bioinfo/Rhaddo/maps.html>]
  15. Scheurle D, DeYoung MP, Binniger DM, Page H, Jahanzeb M, Narayanan R: **Cancer gene discovery using digital differential display.** *Cancer Res* 2000, **60**:4037-43.
  16. Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer PS: **Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks.** *Nat Med* 2001, **7**:673-679.
  17. Khan J, Bittner ML, Saal LH, Teichmann U, Azorsa DO, Gooden GC, Pavan WJ, Trent JM, Meltzer PS: **cDNA microarrays detect activation of a myogenic transcription program by the PAX3-FKHR fusion oncogene.** *Proc Natl Acad Sci U S A* 1999, **96**:13264-13269.
  18. Bortoluzzi S, Rampoldi L, Simonati B, Zimbello R, Barbon A, d'Alessi F, Tiso N, Pallavicini A, Toppo S, Cannata N, Valle G, Lanfranchi G, Danielli GA: **A comprehensive, high-resolution genomic transcript map of human skeletal muscle.** *Genome Res* 1998, **8**:817-825.
  19. Ko MS, Threat TA, Wang X, Horton JH, Cui Y, Wang X, Pryor E, Paris J, Wells-Smith J, Kitchen JR, Rowe LB, Eppig J, Satoh T, Brant L, Fujiwara H, Yotsumoto S, Nakashima H: **Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome.** *Hum Mol Genet* 1998, **7**:1967-1978.
  20. Caron H, van Schaik B, van der Mee M, Baas F, Riggins G, van Sluis P, Hermus MC, van Asperen R, Boon K, Voute PA, Heisterkamp S, van Kampen A, Versteeg R: **The human transcriptome map: clustering of highly expressed genes in chromosomal domains.** *Science* 2001, **291**:1289-1292.
  21. Fujii T, Dracheva T, Player A, Chacko S, Clifford R, Strausberg RL, Buetow K, Azumi N, Travis WD, Jen J: **A preliminary transcriptome map of non-small cell lung cancer.** *Cancer Res* 2002, **62**:3340-6.
  22. Gordon A, McManus A, Anderson J, Fisher C, Abe S, Nojima T, Pritchard-Jones K, Shipley J: **Chromosomal imbalances in pleomorphic rhabdomyosarcomas and identification of the alveolar rhabdomyosarcoma-associated PAX3-FOXO1A fusion gene in one case.** *Cancer Genet Cytogenet* 2003, **140**:73-77.
  23. Bridge JA, Liu J, Qualman SJ, Suijkerbuijk R, Wenger G, Zhang J, Wan X, Baker KS, Sorensen P, Barr FG: **Genomic gains and losses are similar in genetic and histologic subsets of rhabdomyosarcoma, whereas amplification predominates in embryonal with anaplasia and alveolar subtypes.** *Genes Chromosomes Cancer* 2002, **33**:310-21.
  24. Pandita A, Zielenska M, Thorner P, Bayani J, Godbout R, Greenberg M, Squire JA: **Application of comparative genomic hybridization, spectral karyotyping, and microarray analysis in the identification of subtype-specific patterns of genomic changes in rhabdomyosarcoma.** *Neoplasia* 1999, **1**:262-275.
  25. **UniGene** [<http://www.ncbi.nlm.nih.gov/UniGene/lbrowse.cgi?ORG=Hs>]
  26. **Blat at UCSC** [<http://genome.ucsc.edu/cgi-bin/hgBlat.com?mand=start>]
  27. Pruitt KD, Maglott DR: **RefSeq and LocusLink: NCBI gene-centered resources.** *Nucleic Acids Res* 2001, **29**:137-40.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

