Genetic and genomic analysis modeling of germline c-MYC overexpression and cancer susceptibility

Xavier Solé1, Pilar Hernández1, Miguel López de Heredia2,3, Lluís Armengol4, Benjamín Rodríguez-Santiago5,6, Laia Gómez1, Christopher A Maxwell1, Fernando Aguiló7, Enric Condom8, Jesús Abril2, Luis Pérez-Jurado5,6,9, Xavier Estivill4, Virginia Nunes2,3,10, Gabriel Capellá1, Stephen B Gruber11, Víctor Moreno*1 and Miguel Angel Pujana*1

Address: 1Bioinformatics and Biostatistics Unit, and Translational Research Laboratory, Catalan Institute of Oncology, IDIBELL, L’Hospitalet, Barcelona, Spain, 2Medical and Molecular Genetics Center, IDIBELL, L’Hospitalitat, Barcelona, Spain, 3CIBERER-U730, L’Hospitalitat, Barcelona, Spain, 4Genes and Disease Program, Center for Genomic Regulation, Barcelona, Spain, 5Genetics Unit, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain, 6CIBERER-U735, Barcelona, Spain, 7Department of Urology, Bellvitge Hospital University, IDIBELL, L’Hospitalitat, Barcelona, Spain, 8Department of Pathology, Bellvitge Hospital University, IDIBELL, L’Hospitalitat, Barcelona, Spain, 9Program in Molecular Medicine and Genetics, Vall d’Hebron University Hospital, Barcelona, Spain, 10Genetic Unit, Department of Physiology II, University of Barcelona, Spain and 11Departments of Epidemiology, Internal Medicine and Human Genetics, University of Michigan, Ann Arbor, Michigan, USA

Email: Xavier Solé - x.sole@iconcologia.net; Pilar Hernández - phgutierrez@iconcologia.net; Miguel López de Heredia - mlopezheredia@idibell.org; Lluís Armengol - lluis.armengol@crg.es; Benjamín Rodríguez-Santiago - benjamin.rodriguez@upf.edu; Laia Gómez - lgomez@iconcologia.net; Christopher A Maxwell - cmaxwell@iconcologia.net; Fernando Aguiló - faguilo@csub.scs.es; Enric Condom - ecm@csub.scs.es; Jesús Abril - jabril@idibell.org; Luis Pérez-Jurado - luis.perez@upf.edu; Xavier Estivill - xavier.estivill@crg.es; Virginia Nunes - vnunes@idibell.org; Gabriel Capellá - gcapella@iconcologia.net; Stephen B Gruber - sgruber@med.umich.edu; Víctor Moreno* - v.moreno@iconcologia.net; Miguel Angel Pujana* - mapujana@iconcologia.net

* Corresponding authors

Published: 11 January 2008
This article is available from: http://www.biomedcentral.com/1471-2164/9/12

© 2008 Solé et al; licensee BioMed Central Ltd.

Abstract

Background: Germline genetic variation is associated with the differential expression of many human genes. The phenotypic effects of this type of variation may be important when considering susceptibility to common genetic diseases. Three regions at 8q24 have recently been identified to independently confer risk of prostate cancer. Variation at 8q24 has also recently been associated with risk of breast and colorectal cancer. However, none of the risk variants map at or relatively close to known genes, with c-MYC mapping a few hundred kilobases distally.

Results: This study identifies cis-regulators of germline c-MYC expression in immortalized lymphocytes of HapMap individuals. Quantitative analysis of c-MYC expression in normal prostate tissues suggests an association between overexpression and variants in Region 1 of prostate cancer risk. Somatic c-MYC overexpression correlates with prostate cancer progression and more aggressive tumor forms, which was also a pathological variable associated with Region 1. Expression profiling analysis and modeling of transcriptional regulatory networks predicts a functional association between MYC and the prostate tumor suppressor KLF6. Analysis of MYC/Myc-driven cell transformation and tumorigenesis substantiates a model in which MYC overexpression...
promotes transformation by down-regulating KLF6. In this model, a feedback loop through E-cadherin down-regulation causes further transactivation of c-MYC.

**Conclusion:** This study proposes that variation at putative 8q24 cis-regulator(s) of transcription can significantly alter germline c-MYC expression levels and, thus, contribute to prostate cancer susceptibility by down-regulating the prostate tumor suppressor KLF6 gene.

**Background**
Risk of human cancer associated with genetic variation at chromosome 8q24 was first described for prostate cancer in individuals with European ancestry and in African Americans (Risk Region 1) [1,2]. This association was stronger for more aggressive tumor forms [2-4] and for earlier age at diagnosis in African Americans [1,5]. Differences in allele prevalences could account for the higher incidence of prostate cancer in particular populations such as African-Americans [1,2,5]. Subsequently, 8q24 has been associated with risk of prostate cancer by two extra independent regions [6-8] and in risk of breast and colorectal cancer by variation partially overlapping with prostate cancer risk [9-13]. In particular, Haiman et al. [12] first noted the existence of common risk variants for breast and colorectal cancer at 8q24. These observations suggest that multiple cancer genes may exist at 8q24 or, alternatively, that risk variants converge on a common biological mechanism [7].

In these studies risk variants did not map to known genes, with few ESTs identified in relatively close proximity. A proposed mechanism includes differences in genomic structure that would make the 8q24 region more prone to subsequent somatic amplification [14]. The c-MYC gene is of particular interest in this region because its ectopic expression has been shown to induce prostatic neoplasia [15-17]. Here, we analyze genetic and genomic data to provide evidence of 8q24 cis-regulator(s) of germline c-MYC transcription. In addition, genomic data modeling predicts a molecular mechanism linking germline c-MYC overexpression and prostate tumorigenesis.

**Results**

**Genetic association scan for germline expression differences**
Scanning associations between genetic variation at 8q24 and c-MYC gene expression levels in immortalized lymphocytes of HapMap CEU (Utah residents with ancestry from Northern and Western Europe) and YRI (Yoruba in Ibadan Nigeria) individuals showed the existence of clusters of SNPs with nominal \( P \) values < 0.05 (Fig. 1). To assess clustering significance, we examined the proportion of significant SNPs in genomic windows 2- or 4-fold the average size of linkage disequilibrium blocks in CEUs or YRIs, respectively (~42 kb corresponding to ~66 SNPs in CEUs and ~36 kb corresponding to ~61 SNPs in YRIs).

Twenty thousand permutations were performed to evaluate the significance of the observed clustering. One genomic region in CEUs and three regions in YRIs were identified with high density of significant SNPs (Fig. 1).

Variation at the c-MYC locus was observed with a trend in CEUs, which might suggest the existence of cis-regulators in the gene structural elements (blue bar in Fig. 1A). Two variants in this region (rs4645943 C and rs16902364 A) are associated with germline differential expression of c-MYC. The allele frequencies of these SNPs were reported to differ between prostate cancer cases and controls in different populations (i.e. 87.7% (cases) and 77.7% (controls) in Hawaiians; 96.3% (cases) and 95.1% (controls) in CEUs for rs4645943 C) [7]. This observation warrants further genetic analysis of the region with regard to prostate cancer risk.

The scan revealed a possible association between variants in Region 1 of prostate cancer risk and differential germline expression of c-MYC (Fig. 1B). Several significant SNPs within this region were identified: the most significant were rs7387447, rs10808558 and rs16902176 (\( P \) values < 0.01). The rs10808558 A allele showed an association with c-MYC overexpression in YRIs (expression difference of 0.23 log_2 units, 95% confidence interval (CI) 0.06 – 0.41; \( P = 0.007 \)) and this SNP is in low linkage disequilibrium (LD) with the prostate cancer risk variant rs1447295 (\( r^2 = 0.19 \)). Overall, the scan analysis suggests the existence of 8q24 cis-regulators of germline c-MYC transcription in lymphocytes, partially overlapping with Region 1 of prostate cancer risk.

**Expression differences in normal prostate tissues**
Given the possible association of Region 1 variants with germline c-MYC overexpression in immortalized lymphocytes of HapMap individuals, we next examined expression differences in normal prostate tissues. For this analysis we used 54 previously characterized normal prostate tissue samples [18,19] and a real-time qRT-PCR protocol developed for prostate samples [20-22]. Genotyping the prostate cancer-associated rs1447295 variant in these samples identified six heterozygotes harboring the risk allele A (CA genotypes). No significant age differences were found between donors harboring the two different genotypes (CA versus CC; no AA homozygotes were identified). Quantitative RT-PCR study using three gene refer-
Genetic association scan for germline c-MYC differential expression in CEUs and YRIs. (A) Top panel shows results for individual SNPs and bottom panel shows results for significant SNP density in genomic windows of ~42 kb/66 SNPs in CEUs. The red horizontal dashed line marks the nominal $P$ value of 0.05. Variants associated with risk of breast [9], colorectal [10-13] or prostate [1-8, 24] cancer are marked with dashed lines as indicated in the inset. (B) Top panel shows results for individual SNPs and bottom panel shows results for significant SNP density in genomic windows of ~36 kb/61 SNPs in YRIs. Linkage disequilibrium ($D'/LOD$) plots are shown at the bottom for YRIs. Region 1 of prostate cancer risk is shown.
ences (18S, ALAS1 and TBP) identified significant c-MYC overexpression in samples harboring the risk allele relative to CC homozygotes (n = 26) (Wilcoxon rank sum test \( P = 0.004 \)) (Fig. 2A). In addition, no evidence of allele-specific amplification in tumors arising in CA individuals was observed (not shown). These results suggest the involvement of germline c-MYC overexpression in prostate cancer susceptibility.

**Germline copy number variants**

As a possible mechanism explaining germline overexpression, we next examined copy number variants (CNVs) at the c-MYC locus in CEUs and YRIs, and in 322 unrelated individuals from the Spanish general population using a multiplex ligation-dependent probe amplification (MLPA) assay. This assay identified genomic gains at the c-MYC locus at a relatively low frequency in the Spanish general population (< 1%; 2/322) (Additional file 1). However, analysis of rs1447295 genotypes in these individuals did not reveal association with the risk allele and, importantly, none of the CEUs or YRIs showed CNVs with this assay. Therefore, a CNV including c-MYC does not seem to be a major contributor to the risk of prostate cancer and germline c-MYC overexpression associated with Region 1. Wong et al. [23] previously described a CNV including c-MYC but only with genomic losses. This observation corroborates the structural complexity of 8q24 and opens the possibility that different genomic configurations are associated with risk alleles in Region 1 or other 8q24 regions.

**Gene expression analysis in prostate tumors**

Since Region 1 variants were associated with earlier age at diagnosis and high Gleason scores or aggressive tumor forms [1-8,24], we examined the expression level of 8q24 genes in primary prostate tumors and their association with clinical and pathological variables. For these analyses, we used a publicly available expression data set containing different prostate cellular populations isolated using laser-capture microdissection [25].

Comparison of normal versus neoplastic samples showed differential expression of c-MYC (Fig. 2B). Specifically, overexpression appears in the more advanced stages of tumorigenesis such as carcinomas and hormone-refractory metastases (\( t \)-test \( P < 10^{-3} \)). Tomlins et al. [25] previously noted the identification of c-MYC in an "overexpressed in progression" signature. The FAM84B gene at 8q24 also shows overexpression but mainly at earlier stages (\( P = 0.043 \) and \( P = 0.002 \) for intraepithelial neoplasia and carcinomas, respectively), which suggests that FAM84B could also be a target of 8q24 somatic amplification. Analysis of Gleason scores showed a trend for c-MYC overexpression (ANOVA test \( P = 0.056 \)) (Fig. 2C). Association between c-MYC overexpression and high-grade prostate tumors was previously noted by Buttyan et al. [26] and Fleming et al. [27]. These observations point to a causal relationship between somatic c-MYC overexpression and the more aggressive forms of prostate tumors.

**Expression profiles and modeling of transcriptional regulatory networks**

Transcriptional targets of MYC include many genes that were identified as conferring risk of prostate cancer and/or being somatically mutated in prostate tumors [28,29]. We sought to identify which of these genes, particularly those conferring risk of prostate cancer, could be functionally associated with c-MYC by examining the similarity between expression profiles using a data set containing 50 normal tissues and 52 prostate tumors [30]. This analysis revealed strong correlations between c-MYC and the prostate tumor suppressor Kruppel-like factor 6 (KLF6) gene (Fig. 3A). Correlations were positive for c-MYC microarray probes 1973s_at and 37724s_at (Pearson’s correlation coefficient (PCC) = 0.65; \( P < 10^{-13} \)) and negative for 1827s_at (PCC = -0.71; \( P < 10^{-15} \)). Extensive alternative splicing of the c-MYC mRNA could account for this difference [31].

To determine the molecular consequence of the predicted MYC-KLF6 functional association, we generated models of transcriptional regulatory networks in prostate tissues. Using the ARACNe algorithm [32] and the 102 hybridizations of Singh et al. [30], we identified 88 and 111 putative transcriptional targets of MYC and KLF6 in this cell type, respectively (Fig. 3B). The intersection of these two sets contains 25 genes, which is a much larger number of genes than randomly expected using simulations of equivalent gene sets (empirical \( P < 0.001 \)). Importantly, 16 of these genes contain MYC binding sites at their promoters based on TRANSFAC (eukaryotes transcription factors database) matrices [33]. In addition, many known MYC targets [29] were also identified: 22 out of 88 (25%) and 23 out of 111 (20%) of the MYC and KLF6 predicted transcriptional targets, respectively (Fig. 3B). Notably, c-MYC and KLF6 were also directly connected and the KLF6 promoter contains three predicted binding sites for MYC (not shown). A 5-gene recurrence predictor of prostate cancer [34] contains KLF6, three common ARACNe-based predictions between MYC and KLF6 (FOS, JUNB and ZFP36), and PPFIA3, which is functionally related to another predicted target of KLF6 (PPFIBP2) (Fig. 3B). These observations further support the role of KLF6, c-MYC and the ARACNe-based predictions in prostate tumorigenesis.

Comparison of ARACNe-based predictions with the Tomlins et al. data set [25] identified 13 of the 88 predicted MYC transcriptional targets differentially expressed between normal prostate tissues and androgen-independent metastases (FDR-adjusted \( P \) values < 0.05). In addi-
Figure 2
Analysis of c-MYC expression in normal and prostate cancer tissues. (A) Relative expression differences of c-MYC calculated using three gene references with the following formula: \( R = F_{\text{c-MYC}} \times (F_{\text{FTP}} - F_{\text{ALAS1}}) \) where \( F_{\text{gene}} = C_{\text{gene}} - C_{\text{18S}} \). (B) c-MYC expression in prostate cancer progression. Mean expression values are marked by a red solid rhombus. (C) c-MYC expression association study with Gleason scores.
Figure 3
Expression profiling and modeling of transcriptional regulatory networks. (A) Transcriptional profiles of c-MYC and KLF6 in prostate tissues [30] using U95A Affymetrix probes shown in the inset. (B) Integrated transcriptional regulatory networks of MYC and KLF6. Gene function assignment based on GO term annotations and known MYC transcription targets are shown as indicated in the inset.
tion, 20 of the predicted targets were found to be differentially expressed between normal prostate tissues and adenocarcinomas. Notably, ~46–40% of these genes (6/13 and 8/20) were also predicted to be direct transcriptional targets of KLF6 by the ARACNe algorithm, which endorses the putative functional association between MYC and KLF6.

**Analysis of MYC/Myc-driven cellular transformation and tumorigenesis**

To evaluate the functional significance of the predicted shared MYC/KLF6 transcriptional targets, we examined expression data derived from a model of MYC-driven cellular transformation of quiescent human mammary epithelial cells and from MMTV-Myc-driven mammary tumors in mice [35,36]. Of the 25 predicted common targets, 16 (64%) were found to be differentially expressed in cell transformation of quiescent human mammary epithelial cells (Fig. 4A). This proportion is ~2-fold higher than expected by chance taking into account all genes examined in the microarray platform ($\chi^2$-test $P = 0.004$), which substantiates the identification of true MYC targets. Moreover, 11 of the 16 genes contain MYC binding sites at their promoters. Importantly, KLF6 was also identified and showed strong down-regulation in this model ($t$-test $P$ values $< 10^{-3}$) (Fig. 4A).

Analysis of MMTV-Myc-driven mammary tumors in mice showed consistent results with the analysis of quiescent human mammary epithelial cells. Twelve differentially expressed genes were detected, eight of which coincided with the human genes mentioned above (Fig. 4B). Genes that did not overlap between the two studies showed similar trends, for example the human TGIF showed a trend for down-regulation ($P = 0.067$) while it was identified as significant in the study of mice tumors ($P = 0.007$). Importantly, this analysis also revealed Klf6 down-regulation ($P = 0.003$) (Fig. 4B). Overall, the discovery of KLF6/Klf6 down-regulation in two different models of MYC/Myc-driven cell transformation supports the hypothesis that c-MYC germline overexpression could act as a risk factor for prostate cancer by converging on a molecular mechanism such as the functional inactivation of the KLF6 gene or gene product.

Using the MYC/Myc-driven cell transformation models, we next examined the differential expression of known KLF6 transcriptional targets of relevance to epithelial cancers, E-cadherin (CDH1 gene) [37] and p21 (CDKN1A) [38]. This analysis revealed strong down-regulation of CDH1 in the transformation of quiescent human mammary epithelial cells ($P$ values $< 10^{-3}$) and a trend in the model of Myc-driven mouse tumorigenesis ($P = 0.088$). No significant differences were appreciable for CDKN1A or Cdkn1a. These observations suggest that KLF6 down-regulation mediated by germline MYC overexpression could promote epithelial neoplasia by down-regulating E-cadherin.

**Discussion**

Combined analysis of genetic and expression data facilitates the identification of transcriptional regulators acting in any part of the genome [39,40]. Examination of different ethnic groups reinforces the identification of these regulators but also reveals differences between populations [41,42]. Due to their functional and structural complexity, transcriptional regulators are largely undercharacterized. However, it is thought that their genetic variability may be relevant when considering susceptibility to common diseases. Specifically, their causal relationship to cancer is almost unknown since most genetic analyses have been focused on coding regions. Insights into differential germline gene expression and tumorigenesis have been gained mainly from mice models, such as the overexpression of the RAS family of genes [43], Mad2 [44] or c-MYC [45,46].

This study analyzed the hypothesis that variation at 8q24 cis-regulator(s) of transcription could significantly alter germline c-MYC expression levels and, thus, contribute to cancer susceptibility. Although the genetic scanning analysis performed is susceptible to false positives, the existence of true cis-regulator(s) is suggested by the identification of clusters of significant SNPs. Although larger sample series are required to draw definitive conclusions, the quantitative analysis of geneexpression in normal prostate tissues supports the model of c-MYC overexpression associated with Region 1 of prostate cancer risk. Tissue-specific cis-regulator(s) that correlate with additional cancer risk regions at 8q24 may also exist. In a recent study it was noted that tissue specificity is a critical factor in the transcriptional responsiveness of MYC targets [47].

The 8q24 region appears amplified in up to 50% of prostate tumors and c-MYC is thought to be the primary target of these amplifications since it is overexpressed in prostate hyperplasia and neoplasia [25]. Ectopic overexpression of c-MYC/c-Myc is sufficient to immortalize human prostate epithelial cells [17] and has been shown to generate human-like prostate tumors in mice [16]. In addition, c-MYC overexpression in prostate cancer cells enables androgen-independent growth [48]. These observations lead to suggestions of a dual role for c-MYC in prostate cancer. At early stages it would promote proliferation while at later stages it would facilitate androgen-independent growth [17]. Our study further proposes that germline c-MYC overexpression may promote cellular transformation of the normal epithelium and, by extension, risk of prostate cancer by down-regulating the prostate tumor suppressor KLF6 gene. This model is
Figure 4
Expression analysis of predicted MYC/KLF6 transcriptional targets in MYC/Myc-driven cell transformation and tumorigenesis. (A) Results of the analysis of quiescent human mammary epithelial cells [36]. (B) Results of the analysis of MMTV-Myc-driven tumors in mice [35]. Genes (red, up-regulated; green, down-regulated), corresponding microarray probes and two-tailed t-test P values are shown.
hypothetical and mainly based on the application of the ARACNe algorithm, which achieves a reasonable tradeoff between true- and false-positive rates by eliminating the majority of indirect interactions inferred from gene co-expression [49,50]. Experimental corroboration of the predictions generated in this study is therefore needed, particularly in prostate tissues or cell lines.

The KLF6 gene is inactivated in prostate cancer by loss of heterozygosy and/or by somatic mutations identified in tumors, cell lines and xenografts [51]. Recent evidence has extended the role of KLF6 inactivation to several other neoplastic processes as esophageal carcinomas [52], glioblastomas [53], head and neck squamous cell carcinomas [54], hepatocellular carcinomas [55], non-small cell lung cancer [56], ovarian carcinomas [57] and particularly, with regard to 8q24 risk variants, to colorectal cancer [58]. A key KLF6 transcriptional target for epithelial neoplasia is E-cadherin (CDH1 gene), which is a suppressor of cellular invasion [37]. KLF6 directly transactivates the CDH1 promoter resulting in increased levels of its gene product [37]. CDH1 is genetically inactivated in many human cancers and shows reduced or absent expression in approximately 50% of prostate tumors [59], playing a critical role in the transition from a noninvasive to an invasive phenotype [60]. Notably, it has recently been proposed that EphB receptors act as tumor suppressors of colorectal cancer, and possibly breast and prostate cancer, through an E-cadherin-mediated mechanism that compartmentalizes tumor cells in the initial stages of tumorigenesis [61]. Loss of E-cadherin can result in β-catenin nuclear localization and, as a result, the up-regulation of β-catenin

Conclusion

This study proposes that variation at putative 8q24 cis-regulator(s) of transcription can significantly alter germline c-MYC expression levels and, Thus, contribute to prostate cancer susceptibility by down-regulating the prostate tumor suppressor KLF6 gene. We propose a transcriptional regulatory model perturbed in human cancer with a feedback loop for c-MYC.

Methods

Genetic association analysis

We analyzed HapMap genotypes and paired expression data recently made available for immortalized lymphocytes from four ethnic groups and including 210 independent individuals in total (60 Utah residents with ancestry from northern and western Europe; 45 Han Chinese in Beijing; 45 Japanese in Tokyo; and 60 Yoruba in Ilbadan Nigeria; Gene Expression Omnibus (GEO) record GSE6536) [42]. Transcriptional differences were scanned between the 128 and 129 Mb of chromosome 8, corresponding to ~1,530 SNPs (NCBI build 35). Scans were performed in R with the SNPassoc package [68]. The logarithmic effects of alleles were examined. Association of genotypes with the variable response (gene expression level) was calculated by fitting linear equations and P values obtained by assessing the change in deviance against the null model. Association analysis between genotypes, downloaded from the HapMap data release 21a, and gene expression levels were performed using the web-software SNPStats [69]. The D'/LOD plots were generated using the Haplovew software [70].

Microarray gene expression analysis

Using the HapMap lymphocyte expression data [42] and the prostate cancer data of Tomlins et al. [25], matrix series were downloaded from GEO references GSE6536 and GSE6099, respectively. Using the Singh et al. [30] raw data, background correction, normalization and averaging of expression values were performed with the robust multi-
array average (RMA) algorithm. ARACNe Java [49,50] was used to model the gene expression regulatory networks of c-MYC and KLF6. In this analysis, data processing inequality (DPI) tolerance was set to 0.20 and the mutual information (MI) threshold was 0.05. Normalized data sets of MYC/Myc-driven cellular transformation and tumorigenesis were downloaded from the GEO records GSE3151 and GSE3158 [35,36]. Gene probes were matched using the NetAffx (Affymetrix) tool and differentially expressed probes were identified by calculating two-tailed t-test P values.

**Genotyping and quantitative RT-PCR analyses**

Prostate tissue specimens were collected through the Tumor Bank of the Bellvitge University Hospital and the Catalan Institute of Oncology. Genotyping of rs1447295 was performed by direct sequencing of PCR products of genomic DNA using the following forward and reverse primers, respectively: 5'-GAGTTGCACGCCAGACACTA-3' and 5'-TTTCCCATACCCCATTGTA-3'. Quantitative RT-PCR analysis of c-MYC was performed using a protocol previously developed with the LightCycler™ DNA Master SYBR Green I Kit (Roche Applied Sciences) [20-22] and c-MYC primers 5'-CAGCCTGTTAGCCGCTTGATT-3' and 5'-GTGAAATACGCGTCGACCGA-3', and TBP primers 5'-GAACCCAGGCGTATTTTC-3' and 5'-CACACGTCCACCATATT-3'. Relative expression differences were calculated using three gene references (18S, ALAS1 and TBP) with the following formula: \[ R = \frac{F_{c-MYC}}{F_{TBP} \cdot F_{ALAS1}} \]

where \( F_{gene i} \) = \( C_{tgene i} \)- \( C_{18S} \).

**Copy number variant analysis**

MLPA assays were performed following the conventional protocol with 150 ng of DNA, overnight ligation and 32 cycles of PCR. Probes for c-MYC were 5'-GGTTCCTCAACGGGTTGAGGAGGACAGCGTTAACAACGGACT-3' and 5'-TTTTGGCCTGCCTGACCAGATCCCTCTAGATT-GGATCTGCTGGAC-3'.

**Authors’ contributions**

XS participated in the study design, compiled and analyzed the HapMap data, performed the association analysis, and the modeling of transcriptional regulatory networks. PH compiled and analyzed the prostate cancer expression data sets and performed the modeling of transcriptional regulatory networks. FA and EC obtained the tissue specimens. MLH, JA and VN performed the quantitative expression analysis. LA, BRS, LG, LPJ and XE performed the copy number variants analysis. CAM, GC and SBG participated in scientific discussions and helped with the overall interpretation of the data. XS, VM and MAP conceived and designed the study. MAP wrote the original and final versions of the manuscript. All authors read and approved the final version of the manuscript.

**Additional material**

**Additional file 1** MLPA analysis of several cancer loci including c-MYC. Germline genomic gain at c-MYC was identified in a sample (bottom) by comparing relative peak intensities.

Click here for file [http://www.biomedcentral.com/content/supplementary/1471-2164-9-12-S1.png](http://www.biomedcentral.com/content/supplementary/1471-2164-9-12-S1.png)

**Acknowledgements**

The authors are indebted to all those who provided publicly available raw data used in this contribution. We thank Laura González for technical assistance and to three anonymous reviewers for their helpful criticism. Tissue samples were collected through the Tumor Bank of the Bellvitge University Hospital and the Catalan Institute of Oncology, supported by the Tumor Bank Program and the RTICCC CO3/10. This work was also supported by the Catalan Institute of Oncology, the “la Caixa” Foundation (BM05-254-00), the ISCIII (FIS-Pi060545, RCESP-C03/09 and RTICCC-CO3/10) and the Spanish Ministry of Education and Science (SAF-2003/5821 and SAF-2005/00166). CAM is supported by a Beatriz de Pinós fellowship from the Agència de Gestió d’Ajuts Universitaris i de Recerca. MAP is a Ramón y Cajal Researcher with the Spanish Ministry of Education and Science.

**References**

1. Freedman ML, Haiman CA, Patterson N, McDonald GJ, Tandon A, Waliszewska A, Penney K, Steen RG, Ardlie K, John EM, Oakley-Girvan I, Whitemore AS, Cooney KA, Ingles SA, Altshuler D, Henderson BE, Reich D: Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc Natl Acad Sci U S A 2006, 103(38):14068-14073.
2. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, Sigurdsson A, Benediktsson KT, Cajzer JB, Sainz J, Jakobssdottir M, Kostic J, Magnusdottir DN, Ghosh S, Agnarsson K, Birtsgisodottir B, Le Roux L, Olausdottir A, Blondal T, Andredottir M, Gretarsdottir OS, Berthorsson JT, Gudjardottar J, Gylfason F, Thörleifsdottir G, Mañolescu A, Kristjanson K, Geirsson G, Isaksson H, Douglas J, Johannesson JE, Balter K, Wiklund F, Montie J, Y Xu, Suarez BK, Ober C, Cooney KA, Gronberg H, Catalona WJ, Einarsrson GV, Barkardottar RB, Gulcher J, Kong A, Thorsteindottir U, Stefansson K. A common variant associated with prostate cancer in European and African populations. Nat Genet 2006, 38(6):652-658.
3. Suuriniemi M, Agalliu I, Schaid DJ, Johanneson B, McDonnell SK, Iwasaki L, Stanford JL, Osrander EA: Confirmation of a positive association between prostate cancer risk locus and a chromosome 8q24. Cancer Epidemiol Biomarkers Prev 2007, 16(4):809-814.
4. Wang L, McDonnell SK, Slusser JP, Hebbbring SJ, Cunningham JM, Jacobsen SJ, Cerhan JR, Blue ML, Schaid DJ, Thibodeau SN: Two common chromosome 8q24 variants are associated with increased risk for prostate cancer. Cancer Res 2007, 67(7):2944-2950.
5. Schumacher FR, Feigelson HS, Haiman CA, Albanes D, Buring J, Calle EE, Chanock SJ, Colditz GA, Diver WR, Dunning AM, Freedman ML, Ganzano JM, Giovannucci E, Hankinson SE, Hayes RB, Henderson BE, Hoover RN, Kaaks R, Key T, Kolonel LN, Kraft P, Le Marchand L, Ma J, Pike MC, Riboli E, Stampfer MJ, Stram DO, Thomas G, Thun MJ, Travis R, Viramino J, Andriole G, Gelmann E, Willett WC, Hunter DJ: A common 8q24 variant in prostate and breast cancer from a large nested case-control study. Cancer Res 2007, 67(7):2951-2956.
6. Gudmundsson J, Sulem P, Mañolescu A, Amundadottir LT, Gudjardottarson D, Helgason A, Rafnar T, Berghorsthon JT, Agnarsson BA, Baker A, Sigurdsson A, Benediktsson KT, Jakobssdottar M, Xu J, Blondal T,
Kostic J, Sun J, Ghosh S, Stacey SN, Moyer M, Saemundsdottir J, Backman VM, Kristjansson K, Tres A, Partin AW, Albers-Akerr MM, Godin-Durcan T, Walsh PC, Swinkels DW, Navarrete S, Isachsen SD, Aben KK, Graaf T, Cashy J, Ruiz-Echarri M, Willey KE, Suarez BK, Witjes JA, Frigge M, Ober C, Jonsson E, Einarsdotir GV, Mayordomo JL, Kieneyen LA, Isacs WB, Catalona TJ, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kangel A, Stefansson K: Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 2007, 39(5):631-637.

7. Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewski SP, Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewski SP: Comprehensive analysis of common copy-number variation in the human genome. Nat Genet 2007, 39(5):645-649.

8. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello M, Feinleib M, Hong WK, Wu AH, Reich D, Henderson BE: Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet 2007, 39(5):638-644.

9. Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, Swerdlow PJ, Morrison J, Field H, Luben R, Wareham NH, Houlston RS, Healey CS, Bowman R, Armstrong BK, Claussen LC, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gabo-Novoselova A, Wang C,NC, Sozzi G, Negro F, Mapp K, Chanock SJ, Rebbeck TR, Ponder BAJ: A common genetic risk variant of colorectal cancer at 8q24.21. Nat Genet 2007, 39(1):41-49.

10. Thirunavukkarasu S, Savarese A, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, Martin L, Sellig S, Jedrzejczyk AM, Amundsen C, Grondahl N, Breivik S, Neudt W, Hansen K, Anfinsen A, Buecher B, Riboli E, Kury S, Chanock SJ, Rebbeck TR, Ponder BAJ: Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007, 447(7148):1087-1093.

11. Tanzie GW, Greenwood CMT, Rangrelli J, Kustra R, Tenesa A, Farrington SM, Prendergast GS, Olschwang S, Chiang T, Crowdy E, Ferretti V, Lafamme P, Sundararajan S, Roumy S, Olivier JF, Robidoux F, Sladek R, Montpetit A, Campbell P, Bezieux S, O’Shea AM, Zogopoulos G, Cotterchio M, Newcomb P, McClaughlin J, Younghusband B, Green R, Green J, Porteous MEM, Campbell B, Hlante S, Sahatçiu Oktay S, Gray R, Thomas H, Petasnick JP, Shared Genome-wide association studies of colorectal cancer. Nat Genet 2007, 39(7):948-954.

12. Sato K, Qian J, Sleazak JM, Lieber MM, Bostwick DG, Bergstralh EJ, Jenkins RB: Clinical significance of alterations of chromosome 8 in high-grade, advanced, nonmetastatic prostate carcinoma. J Natl Cancer Inst 1999, 91(18):1574-1580.

13. Thirunavukkarasu S, Kurebayashi K, Fernandez S, Stenestrand U, Lai YF, Roberts RL, Hayward SW: Unsupervised c-MYC expression in benign prostatic epithelium causes a cancer phenotype. Prostate 2005, 63(4):369-384.

14. Elwood-Pyke T, Graeber TG, Wongvipat J, Irula-Arispe ML, Zhang J, Markers R, Thomas GV, Sawyer CL: Myc-driven murine prostate cancer shares molecular features with human prostate tumors. Cancer Cell 2003, 4(3):223-238.

15. Gil J, Kerai P, Lleonart M, Bernard D, Cigudosa JC, Peters G, Carnero A, Beach D: Immortalization of primary human prostate epithelial cells by c-Myc. Cancer Res 2005, 65(17):7179-7185.

16. Gomez-Zaera M, Abril J, Gonzalez L, Aguilo F, Condor E, Nadal M, Nunes V: Identification of somatic and germline mitochondrial DNA sequence variants in prostate cancer patients. Mutat Res 2006, 595(1-2):42-51.

17. Linja MJ, Saarela SM, Jukkala CH, Ylikahri R, Kallio VM, Korpi T: Amplification and overexpression of androgen receptor gene in hormone-refractory prostate cancer. Cancer Res 2001, 61(9):3550-3555.

18. Saarela SM, Jukkala CH, Ylikahri R, Kallio VM, Korpi T: Amplification and overexpression of androgen receptor gene in hormone-refractory prostate cancer. Cancer Res 2001, 61(9):3550-3555.

19. Linja MJ, Saramaki OR, Tammela TL, Vessella RL, Visa korpi T: Amplification and overexpression of androgen receptor gene in hormone-refractory prostate cancer. Cancer Res 2001, 61(9):3550-3555.

20. Linja MJ, Saramaki OR, Tammela TL, Chang CT, Brinkman AO, Visakorpi T: Amplification and copy number analysis of TP53, E1F3F3 and MYC genes in breast and prostate cancer. J Cancer Res 2004, 90(5):1041-1046.

21. Wu AH, Reich D, Henderson BE: Comprehensive analysis of common copy-number variation in the human genome. Am J Hum Genet 2007, 80(1):91-104.

22. Severi G, Hayes VM, Padilla EJ, English DR, Southey MC, Sutherland RL, Hopper JL, Giles GG: The common variant rs1447295 on chromosome 8q24 drives prostate cancer risk: results from an Australian population-based case-control study. Cancer Epidemiol Biomarkers Prev 2007, 16(3):610-612.

23. Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, Sundaram S, Wei JT, Rubin MA, Pienta KJ, Shah RB, Chinnaiyan AM: Integrative genomic and molecular feature modeling of prostate cancer progression. Nat Genet 2007, 39(1):41-51.

24. Buttayn R, Sawczuk IS, Benson MC, Siegal JD, Olsson CA: Expression of the c-myc protooncogene in high-grade human prostate cancers. Prostate 1987, 11(4):327-337.

25. Fleming VH, Hamel A, MacDonald R, Ramsey E, Pettigrew NM, Johnston B, Dodd JG, Matsuk J: Expression of the c-myc protooncogene in human prostatic carcinoma and benign prostatic hyperplasia. Cancer Res 1986, 46(3):1535-1538.

26. Futreal PA, Coin L, Marshall M, Dunn D, Hubbard T, Wooster R, Rahman N, Stratton MR: A census of human cancer genes. Nat Rev Cancer 2004, 4(3):177-183.

27. Zeller KL, Jegga AG, Aronov BJ, O’Donnell KA, Dang CV: An integrated database of genes responsive to the Myc oncopgenic transcription factor: identification of direct genomic targets. Genome Biol 2003, 4(10):R69.

28. Singh D, Feigbo PG, Ross K, Jackson DG, Manola J, Ladd C, Tamayo P, Renshaw AA, D’Amico AV, Richie JP, Lander ES, Loda M, Kantoff PW, Theroux KR, Sellers WR: Gene expression correlates of clinical prostate cancer behavior. Cancer Cell 2002, 1(2):203-213.

29. Bodezetz M, Brison O: Characterization of new human c-myc mRNA species produced by alternative splicing. Gene 1996, 174(1):115-120.

30. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A: Reverse engineering of regulatory networks in human B cells. Nat Genet 2005, 37(4):382-390.

31. Matsys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hor nsicker K, Karas D, Kel AE, Kel-Margoulis OV, Kroos DU, Land S, Lewicky-Potapov B, Michael H, Munch R, Reuter I, Roters S, Saxel H, Scheer M, Thiele S, Wingender E: TRANSFAC: transcriptional
regulation, from patterns to profiles. Nucleic Acids Res 2003, 31(1):374-378.

34. Silinsky GW, Glinsky AB, Stephenson AJ, Hoffman RM, Gerald WL: Gene expression profiling predicts clinical outcome of prostate cancer. J Clin Invest 2004, 113(6):913-923.

35. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, Olson JA Jr, Marks JR, Dressman HK, West M, Nevins JR: Oncogenic pathway signatures as guides to targeted therapies. Nature 2006, 439(7074):353-357.

36. Huang E, Ishida S, Pittman J, Dressman H, Bild A, Kloos M, D’Amico M, Festell RG, West M, Nevins JR: Gene expression phenotypic models that predict the activity of oncogenic pathways. Nat Genet 2003, 34(2):226-230.

37. DiFeo A, Narla G, Camacho-Vanegas O, Nishio H, Rose SL, Buller RE, Bernard D, Pourtier-Manzanedo A, Gil J, Beach DH: Ito G, Uchiyama M, Kondo M, Mori S, Usami N, Maeda O, Kawabe T, Hasegawa Y, Shimokata K, Sekido Y: KLF6 is a novel transcriptional target of the KLF6 tumor suppressor. Oncogene 2006, 25(44):6026-6031.

38. Narla G, Kremer-Tal S, Matsumoto N, Zhao X, Yao S, Kelley K, Tarocchi M, Friedman SL: In vivo regulation of p21 by the Kruppel-like factor 6 tumor-suppressor gene in mouse liver and human hepatocellular carcinoma. Oncogene 2007, 26(30):4428-4434.

39. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG, Spielman RS: The genetics of variation in gene expression. Nat Genet 2002, 32 Suppl:522-525.

40. Buckland PR: Allele-specific gene expression differences in humans. Hum Mol Genet 2004, 13 Spec No 2:R255-60.

41. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG: Common genetic variants account for differences in gene expression among ethnic groups. Nat Genet 2007, 39(2):226-231.

42. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavare S, Deloukas P, Hurles ME, Dermitzakis ET: Relative impact of nucleotide and copy number variation on human gene expression. Science 2007, 315(5813):848-853.

43. Mangues R, Seidman I, Gordon JW, Pellicer A: Reverse engineering cellular networks in a mammalian cell culture. BMC Bioinformatics 2007, 8(Suppl 1):S7.

44. Sotillo R, Hernando E, Diaz-Rodriguez E, Teruya-Feldstein J, Cordon-Cardo C, Levine SW, Benezra R: Involvement of Kruppel-like factor 6 (KLF6) in hepatocellular carcinoma. Hepatology 2004, 39(2):485-495.

45. Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Abecasis GR, Voight BF, Goldstein DB, assaulted by copy number variation. Hum Mol Genet 2006, 15(R3):R255-60.

46. Mukai S, Hiyama T, Tanaka S, Yoshihara M, Arihiro K, Chayama K: SNPStats: an R package to perform whole genome association studies. Bioinformatics 2007, 23(4):464-465.

47. Mukai S, Miyasaka T, Tanaka S, Yoshihara M, Arishi K, Chayama K: Involvement of KLF6 locus in colorectal cancer. Int J Cancer 2007, 121(9):1976-1983.

48. Teixeira MS, Camacho-Vanegas O, Fernandez Y, Narla G, Difeo A, Lee B, Kalir T, Friedman SL, Schlecht NF, Genden EM, Urken M, Brandwein-Gensler M, Martignetti JA: KLF6 allelic loss is associated with tumor recurrence and markedly decreased survival in head and neck squamous cell carcinoma. Int J Cancer 2007, 121(9):1976-1983.

49. Kremer-Tal S, Reyes HL, Narla G, Thung SN, Schwartz M, Difeo A, Katz B, Bruij J, Bouilac-Sage P, Martignetti JA, Friedman SL: Frequent inactivation of the tumor suppressor Kruppel-like factor 6 (KLF6) in hepatocellular carcinoma. Hepatology 2004, 39(5):1047-1052.

50. Ito G, Uchiyama M, Kondo M, Mori S, Usami N, Maeda O, Kawabe T, Hasegawa Y, Shimokata Y, Sekido Y: KLF6 is a novel transcriptional target of the KLF6 tumor suppressor. Oncogene 2007, 26(30):4428-4434.

51. Teixeira MS, Camacho-Vanegas O, Fernandez Y, Narla G, Difeo A, Lee B, Kalir T, Friedman SL, Schlecht NF, Genden EM, Urken M, Brandwein-Gensler M, Martignetti JA: KLF6 allelic loss is associated with tumor recurrence and markedly decreased survival in head and neck squamous cell carcinoma. Int J Cancer 2007, 121(9):1976-1983.