Mutagenesis vol. 27 no. 2 pp. 161–167, 2012 doi:10.1093/mutage/ger057

Gene expression variations: potentialities of master regulator polymorphisms in colorectal cancer risk

Barbara Pardini[1,*], Alessio Naccarati[1], Pavel Vodicka[1,2] and Rajiv Kumar[3]

[1]Department of Molecular Biology of Cancer, Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, Videnska 1083, 14220 Prague 4, Czech Republic, [2]Institute of Biology and Medical Genetics, First Medical Faculty, Charles University, Albertov 4, 12800 Prague 2, Czech Republic and [3]Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany.

*To whom correspondence should be addressed. Institute of Experimental Medicine, Videnska 1083, 14200 Prague 4, Czech Republic. Tel: +420 2 41062694; Fax: +420 2 41062782; Email: pardini@biomed.cas.cz

Received on July 13, 2011; revised on July 29, 2011; accepted on August 01, 2011

Colorectal cancer (CRC) is one of the most common cancers worldwide with a peak of incidence in industrialised countries. It is a complex disease related to environmental and genetic risk factors. Low-penetrance genetic variations contribute significantly to sporadic and familial form of CRC. Genome-wide association studies (GWAS) have uncovered numerous robust associations between common variants and CRC risk; only a few of those were protein altering non-synonymous polymorphisms. One of the hypotheses is that non-coding and intergenic variants may change the expression levels of one or several target genes and, thus, account for a fraction of phenotypic differences, including susceptibility to CRC. Such genetic variations have been detected as expression quantitative loci (eQTLs) that show linkage/association to a large number of genes and have been defined as "master regulators of transcription". In the present work, we overview the potentialities to use results from GWAS and eQTL studies in the identification as well as investigation of master regulators in CRC susceptibility.

Colorectal cancer

General information

Colorectal cancer (CRC) is one of the most common malignancies in the world, with particularly high incidence rates in Western countries (1). Worldwide, there were estimated 1 230 000 new cases in the year 2008. CRC is also the third leading cause of cancer death, with an estimated 680 000 deaths reported in 2008 (2). One of the consistent etiological factors associated with risk of the disease has been the Western-style diet, a highly caloric and rich in animal fat, which in combination with sedentary lifestyle constitutes an extremely high risk factor. Those inferences have been reinforced by epidemiological studies on migrant populations. One of the possible explanations for the correlation between CRC and diet is perhaps direct mucosal contact, in colon and rectum, with food components and probable exposure to diet-induced metabolic and physiologic changes (2–6). However, CRC like other multifactorial diseases is fundamentally driven and modulated by individual genetic backgrounds that most likely interact with environmental risk factors and influence risk at population level.

Colorectal tumours progress through a series of clinical and histopathologic stages that range from single crypt lesions (aberrant crypt foci) through small benign tumours (adenomatous polyps) to malignant cancers (carcinomas). The transition of normal epithelia cells to carcinoma cells has been shown to involve a series of somatic genetic changes that include inactivation of tumour suppressor genes as well as oncogenic activation (7). The major somatic genetic alterations include activating mutations in *K-RAS*, inactivation of tumour suppressor genes on chromosomes 17p (*TP53*), 5q and 18q (8). Mutations in *K-RAS* occur in about 50% of CRC and in adenomas. Screening of *K-RAS* mutations constitutes an integral part of regular diagnosis prior to the treatment of patients with metastasised disease (9). Somatic mutations in adenomatous polyposis coli (*APC*) gene, located on chromosome 5q, are present in over 80% of sporadic CRC. Inherited mutations in the gene are responsible for familial adenomatous polyposis (FAP) coli, a genetic autosomal dominant disease. Somatic alterations in the *APC* gene that occur in sporadic CRC are similar to inherited truncating mutations in FAP. *APC* mutations are the earliest lesion and are, therefore, considered as gatekeeper event in colorectal tumorigenesis (10).

Genetics of CRC

Colorectal carcinoma is generally classified into three categories. Familial CRC (around 30% of total CRC cases) refers to patients with at least one blood relative with CRC or an adenoma but with no specific germline mutations or clear pattern of inheritance. The majority, approximately 67% of CRC cases, are sporadic or show a pattern of familial aggregation, not fitting into models of Mendelian inheritance (10). The remaining 3% of CRC patients belongs to cases with inherited CRC syndromes (11,12). Those are carriers of highly penetrant germline mutations in cancer susceptibility genes. The most representative syndromes are the ones characterised by multiple benign colorectal polyps and others without polyposis. The first group includes several polyposis syndromes, such as FAP, Peutz-Jegher syndrome, familial juvenile polyposis. The hereditary nonpolyposis colorectal cancer, also known as Lynch syndrome, is characterised by germline mutations in different DNA mismatch repair genes (13).

The complex etiology of sporadic CRC involves a combinatorial impact of genetic and environmental risk factors (10,14). Epidemiological studies have identified several risk factors, including meat, tobacco and alcohol consumption, that modulate risk of CRC (reviewed in ref. 5). On the other hand, vegetable consumption, prolonged use of non-steroidal anti-inflammatory drugs, oestrogen replacement therapy and physical activity are known as protective factors (5). Other significant etiological factors are chronic inflammatory bowel

diseases and a positive family history of CRC. Population-based studies showed that ulcerative colitis results in a 20-fold increased incidence of CRC and a 4.4-fold increase in mortality, while Crohn's disease causes 3-fold increase in the risk of colorectal malignancies (15,16). Family history is established to be one of the strongest risk factors for the development of the disease, as almost 30% of the individuals report a positive family history (17).

It had been long hypothesized and subsequently shown experimentally that low-penetrant genetic variants are probably causal for the inter-individual variation in susceptibility to CRC. CRC, as most of forms of human cancers, shows complex inheritance in the general population. Contributors to these complex patterns can be combination of genetic variants with common-to-rare frequency and low-to-intermediate penetrance (18). Although estimated inherited susceptibility accounts for about 35% of variance in CRC risk, high-penetrant germline mutations account for only 6% of cases (19,20). Much of the remaining variation in genetic risk is likely to be a consequence of the co-inheritance of multiple low-penetrance variants, which are common in general population. The ''common-disease common-variant'' model of CRC implied that association analyses based on scans of polymorphic variants should be a powerful strategy for identifying low-penetrance susceptibility alleles (14). While the association studies based on a candidate gene approach did identify some variants involved in CRC susceptibility, but success of such investigations remained limited and many times results were not reproducible by independent studies. Recent meta-analyses of various candidate-gene based case-control studies have provided limited support for the role of variants in the *MTHFR*, *CCND1*, *GSTT1*, *XPC*, *NQO1* and *NAT2* genes (21–27).

GWAS, genetic expression variation and CRC

The advent of so-called genome-wide association studies (GWAS) has resulted in robust identification of genetic variants associated with susceptibility of various complex diseases. Several GWAS have also been applied to investigate the impact of common genetic variations on CRC susceptibility (28–43). So far, various chromosomal loci have been shown to modestly increase CRC risk (Table I) (28–36,41,43). The identified variants, with minor allele frequencies between 7 and 90%, have been shown to increase risk of CRC with allelic odds ratios < 1.3 (44). Though, the individual associated variants identified through GWAS impart only a modest risk of CRC, the population attributable fractions have been shown to be significant because of substantial minor allele frequencies. Moreover, through the possibility of a concerted effect, the variants can potentially influence an individual's risk of developing CRC significantly (14). Based on the location within the genome, most of the identified associated single-nucleotide polymorphisms (SNPs) cannot be *per se* causal but rather indicate the possibility of being in linkage with the ''real'' causal variants. Nevertheless, the role of some of the associated variants in transcription either through *cis*- or *trans*-regulation remains plausible as indicated by several investigators (45–48). In one such study, it has been shown that a variant at chromosome 8q23.3 may influence the transcriptional regulation of eukaryotic translation initiation factor 3H (*EIF3H*), providing support for the functional significance of the variant and observed association with CRC (46). This class

Table I. SNPs associated with risk of CRC identified and validated through GWAS

Polymorphism	Chromosome	Locus/genes (major/minor alleles)[a]	References	Cases/controls[b]	OR[c]	95% CI[d]	*P*-value
rs16892766	8q23.3	*EIF3H* (A/**C**)	Tomlinson *et al.* (36)	922 (17872)/927 (17526)	1.27	1.20–1.34	3×10^{-18}
rs 10795668	10p14	Intergenic (**G**/A)	Tomlinson *et al.* (36)	922 (17872)/927 (17526)	1.12	1.10–1.16	3×10^{-13}
rs6983267	8q24.21	*c-MYC* (**G**/T)	Tomlinson *et al.* (36)	922 (17872)/927 (17526)	1.24	1.17–1.33	7×10^{-11}
			Tomlinson *et al.* (30)	930 (7334)/960 (5246)	1.27	1.16–1.39	1×10^{-14}
			Zanke *et al.* (31)	1257 (6223)/1336 (6443)	1.17	1.12–1.23	3×10^{-11}
			Cui *et al.* (41)	1583 (4809)/1898 (2973)	1.18	1.11–1.25	2×10^{-8}
rs 4939827	18q21.1	*SMAD7* (**T**/C)	Broderick *et al.* (28)	940 (7473)/965 (5984)	1.16	1.09–1.27	1×10^{-12}
			Tenesa *et al.* (35)	981 (16476)/1002 (15351)	1.2	1.16–1.24	8×10^{-28}
			Tomlinson *et al.* (36)	922 (17872)/927 (17526)	1.18	1.10–1.25	2×10^{-6}
rs4779584	15q13.3	*GREM1/SCG5* (C/**T**)	Tomlinson *et al.* (36)	922 (17872)/927 (17526)	1.23	1.14–1.34	5×10^{-7}
rs3802842	11q23.1	*C110rf93* (A/**C**)	Tenesa *et al.* (35)	981 (16476)/1002 (15351)	1.11	1.08–1.15	6×10^{-10}
rs4444235	14q22.2	*BMP4* (T/**C**)	Houlston *et al.* (32)	1902 (4878)/1929 (4914)	1.11	1.08–1.15	8×10^{-10}
rs9929218	16q22.1	*CDH1* (G/**A**)	Houlston *et al.* (32)	1902 (4878)/1929 (4914)	1.10	1.06–1.12	1×10^{-8}
rs10411210	19q13.11	*RHPN2* (**C**/T)	Houlston *et al.* (32)	1902 (4878)/1929 (4914)	1.15	1.10–1.20	5×10^{-9}
rs961253	20p12.3	Intergenic (C/**A**)	Houlston *et al.* (32)	1902 (4878)/1929 (4914)	1.12	1.08–1.16	2×10^{-10}
rs10936599	3q26.2	*MYNN* (**C**/T)	Houlston *et al.* (38)	3334 (14851)/4628 (15569)	1.04	1.04–1.10	3×10^{-8}
rs6687758	1q41	*DUSP10* (A/**G**)	Houlston *et al.* (38)	3334 (14851)/4628 (15569)	1.09	1.06–1.12	2×10^{-9}
rs4925386	20q13.33	*LAMA5* (**C**/T)	Houlston *et al.* (38)	3334 (14851)/4628 (15569)	1.08	1.05–1.10	2×10^{-10}
rs11169552	12q13.12	*LARP4/DIP2* (**C**/T)	Houlston *et al.* (38)	3334 (14851)/4628 (15569)	1.09	1.05–1.11	2×10^{-10}
rs6691170	1q41	*DUSP10* (G/**T**)	Houlston *et al.* (38)	3334 (14851)/4628 (15569)	1.06	1.03–1.09	1×10^{-9}
rs7758229	6q25.3	*SLC22A3* (G/**T**)	Cui *et al.* (41)	1583 (4809)/1898 (2973)	1.28	1.18–1.39	8×10^{-9}
rs7014346	18q24.21	*POU5FIP1/HsG57825*/DQ515897 (G/**A**)	Tenesa *et al.* (35)	981 (16476)/1002 (15351)	1.19	1.15–1.23	9×10^{-26}

[a]The risk allele for each SNP is highlighted in bold. *EIF3H*, eukaryotic translation initiation factor 3, subunit H; *c-MYC*, v-myc avian myelocytomatosis viral oncogene homologue; *SMAD7*, SMAD family member 7; *GREM1*, gremlin 1; *SCG5*, secretogranin V; *C110rf93*, chromosome 11 open reading frame 93; *BMP4*, bone morphogenetic protein 4; *CDH1*, E-cadherin; *RHPN2*, Rho GTPase binding protein 2; *MYNN*, myoneurin; *DUSP10*, dual specificity phosphatase 10; *LAMA5*, laminin alpha 5; *LARP4*, La ribonucleoprotein domain family, member 4; *DIP2*, DIP2 disco-interacting protein 2 homologue A (Drosophila); *SLC22A3*, solute carrier family 22 (extraneuronal monoamine transporter), member 3; *POU5FIP1/HsG57825*/DQ515897, hypothetical LOC727677.
[b]In brackets are numbers of individuals used for validation part.
[c]OR, odds ratio.
[d]95% CI, 95% confidence interval.

of susceptibility variants has been suggested to be of potential public health importance, allowing risk stratification within populations (14). The identification of new risk variants may also be helpful in characterising, more specifically, new cancer pathways that may lead to new prevention or treatment strategies for CRC (14).

Despite remarkable success of GWAS in identifying disease-associated loci, a gnawing gap remains in understanding the mechanism that leads to the associated increased risk. The genetic risk may be probably due to both highly prevalent loci with very modest effect sizes and rare loci with strong effect sizes (49). There have been, so far, only some rare examples in which the observed pattern of genetic association corresponded to a functional candidate variant with a precise localisation of the functional locus. In the majority of GWAS, regional linkage disequilibrium extends across multiple genes and the disease-associated variants serve as proxies for unrecognised non-coding variation, precluding claims of disease-specific gene identification (50). Another common feature characterising GWAS is that many disease-linked SNPs identified to date are located within introns or inter-genic regions of human genome, which have no direct relations to known protein-coding genes or microRNA genes (51). It has been correctly suggested that the biological interpretation of GWAS signal is daunting. Candidate loci fall either in gene deserts or in regions with equally plausible causative genes (52). Thus, non-canonical mechanisms of phenotype-altering effects of genetic variations cannot be ruled out. For example, Glinskii *et al.* proposed that inter-genic DNA sequence variations associated with human disorders may affect phenotypes through *trans* action via non-protein-coding SNP sequence-bearing RNA intermediates (51,53). An alternative approach to solve these problems and to reduce costs associated with GWAS and to overcome limitations of multiple testing corrections and sample size is to select SNPs on basis of a functional effect, such as changes in gene expression levels (52,54–56).

Expression quantitative trait loci studies: a way to mapping DNA variants that influence gene expression

The levels of expression of different genes vary among individuals and contribute to phenotypic diversity and differences in disease susceptibility (57). It has been shown that several chromosomal regions contain germline determinants that regulate gene expression phenotypes (58). The quantitative differences in gene expression might be responsible for a large source of natural variation in populations, as well as crucial in accounting a large fraction of phenotypic variability (59,60). Variable gene expression levels among individuals can be analysed like other quantitative phenotypes (61). Functional genomics, in particular expression quantitative trait loci (eQTL) studies, are useful for determining relationship between sequence variation in human population and phenotypic diversity (62). The main goal of eQTL studies in humans has been the identification of DNA variants (polymorphisms) that influence the expression levels of genes (also defined as gene expression phenotype). The basic idea is that sequence variation may be crucial in affecting the steady-state level of mRNA molecules of a particular gene in a specific cell type or tissue. eQTLs studies have led to interesting results for the identification of regulatory regions and DNA sequence variants that may modulate levels of expression of genes in a range of organisms. Interestingly, eQTL studies could fill in the gaps in understanding the causal reasons for the association of loci with the disease identified through GWAS (Figure 1) (63).

eQTL studies connect variation at DNA level with the quantitative variation in RNA transcription (61). Over 15 million SNPs have been identified so far and while some of those are functional, the majority is most probably neutral (64). eQTL studies simplify the search for functional variants that regulate gene expression through scan of genome for regulators of gene expression without the need for *a priori* knowledge of regulatory mechanisms (61). Gene expression levels are controlled by a combination of *cis*- and *trans*-acting regulators. Gene expression variation may also be considered a possible candidate to establish a link between DNA sequence variability and clinical phenotypes (65). This represents a paradigm shift from the traditional point of view about the role of genes in the susceptibility to human diseases as an effect of their variability in the sequence.

Several disease-linked associations with significant eQTLs have emerged and such studies have also been extended to cancer susceptibility (63). eQTLs and gene expression differences have been evaluated to elucidate the causal relation between validated SNPs identified through GWAS and the risk of lung cancer in never smokers (66). Out of 44 candidate SNPs identified that might alter lung cancer risk in the latter, only one was replicated in four independent studies. From the eQTL analysis, the investigators observed a strong correlation between genotypes of the replicated SNP and the transcription levels of *GPC5* gene in normal lung tissues, showing that genetic variants at a specific locus (13q31.3), associated with susceptibility to lung cancer in never smokers, alter the expression of *GPC5* (66).

Out of 14 loci associated with susceptibility to CRC, discovered through various GWAS, identification of the causal variants and putative functionalities for some have been attempted. In one such study, investigators fine mapped the risk-associated region, identified earlier through GWAS, for identification of functional variants (47). In the three fine-mapped regions, a cluster of SNPs within the narrow areas showed stronger signals for association with the CRC risk than the surrounding GWAS identified SNPs. Using an eQTL browser, some of the top associated variants in the regions analysed have shown an association with different transcript levels (47). Interestingly, some of the genes previously proposed as best candidates in those blocks (*EIF3*, *CDH1* and *RHPN2*) showed no association with different eQTLs. Those genes had been originally selected because of their putative involvement in CRC tumorigenesis and vicinity to the risk-associated tagSNPs. On the other hand, a SNP in an intron of *ZFP90* was found to be associated with CRC risk and with different transcript levels of the gene. Similarly, *UTP23* appeared to be a possible target of the genetic variation associated with CRC in the 8q23.3 region (47). However, it is still difficult to draw too many conclusions for now, since eQTL data sets are not available specifically for colorectal cells and the observed associations have been detected only in blood cells. It is possible that implementation of expression studies in colorectal mucosa could clarify the functionality of the SNPs associated with the CRC risk. In fact, most eQTL studies, currently, are conducted on lymphoblastoid cell lines, which do have some utility for the interpretation of human disease associations, particularly with immunity-related phenotypes (52). At the same time potentiality of artefacts associated with

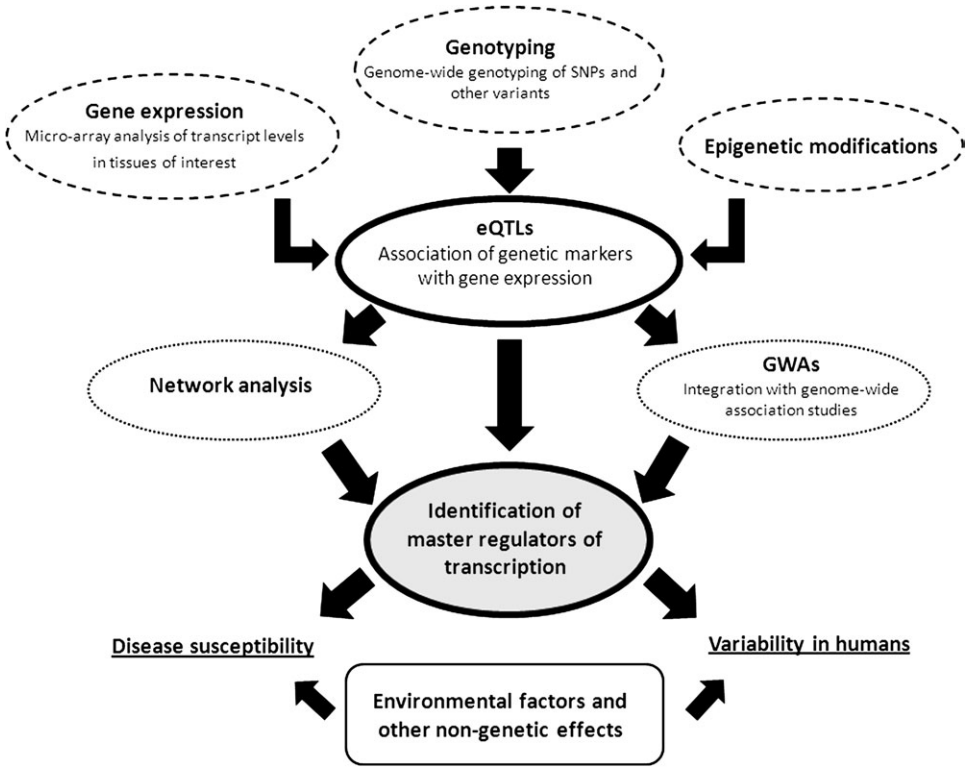

Fig. 1. Integration between GWAS and eQTL mapping in cancer susceptibility studies (modified from ref. 63).

immortalisation, subsequent passage, and growth conditions prior to harvest cannot be precluded (67).

Hotspots or master regulators

It is likely that genetic variation within a gene, encoding a transcription factor with multiple targets, can potentially influence expression of most of its target genes. Such variations are of particular biological interest, which can be, in principle, used to identify groups of genes that are controlled by a common transcription factor (68). The term ''master regulators'' has been applied to those genetic variants that have been detected as eQTL and which show linkage/association to a large number of genes (58). One of the common features observed in studies of transcript expression is the presence of hotspots, that is the individual loci that affect a large number of transcripts (69). Hotspots or ''master regulators of transcription'' are defined as the loci for which the number of associated transcripts exceeds the expected, based on assumption of random distribution along the genetic map (58,70). The description of master regulators is currently not uniform. One approach takes into consideration the number of expression traits that best map to a genetic marker regardless of statistical significance at the whole-genome level (71). Other studies have divided the genome into bins and then counted the number of significant linkages within each bin (58,72–74). Another study divided the autosomal genome of lymphoblastoid cell lines into 491 windows of 5Mb and determined the number of regulators mapping to each window (71). Assuming random distribution across the genome of phenotype regulators, the probability of finding six or more hits per window would be very low. On the contrary, in that study, investigators found two hotspots with six or more hits. In others studies, the number of whole-genome significant linkages at each marker have been counted to identify transcriptional master regulators (70,75,76). In one study, two different linkage strategies to localise master regulators of transcription were applied. One strategy involved classical-variance components with use of information from grandparents and the other based on a linkage method by regression with floating variable selection that allowed testing linkage models with combinations of genetic locations. The former method identified 11 hotspots and the latter 17 with an overlap of five hotspots (71).

Studies in yeast and other organisms have identified hotspots that contain genetic variants that influence multiple expression phenotypes (reviewed in refs. 61 and 77). Human studies, thus far, have shown less uniform results. The presence of master regulatory regions has been either reported (58,78) or not (54,79–82). Surprisingly, only very few verified hotspots have been identified. The majority of the results indicate a large abundance of local eQTLs that coincide with the position of the gene and are presumably *cis*-acting polymorphisms in the promoter region. Genetic variants in hotspots, due to their characteristics, are expected to act more in a *trans* manner than observed. It is likely that the differences among studies are partly because of differences in power to detect *trans*-acting variants. Distal eQTLs are more difficult to be identified. It is probable that indirect regulation mechanism results in reduced statistical power that limits the reliable detection of hotspots (77,83).

The mechanism of gene expression regulation by master regulators, identified thus far, currently can be at the best speculated. The target genes with phenotypes that map to the same hotspots often share similar functions or reside close to each other. As genes that share functions are often co-regulated, their polymorphic regulators would appear as hotspots in eQTL studies. The expression levels of co-

regulated genes frequently show significant correlations that may be biologically important. But sometimes the number of phenotypes mapping to hotspots can be overestimated (84). Despite instances of false positive associations, some of the co-mapped clusters have been experimentally verified. Cyclin H was validated as a new upstream regulator of cellular oxidative phosphorylation as well as a transcriptional regulator of genes comprising a hotspot (85). Similarly, other studies have reported identification of causal regulators and hotspots (51,86,87).

The distant eQTLs and their hotspots, from the reported studies, seem to be scarce and are difficult to identify; those already reported need to be interpreted with caution. The rarity of hotspots may be due to the limited power of the initial studies but could also be attributed to factors that have been hypothesized over the years (77,88,89). In fact, it might be that biological systems are robust against subtle genetic perturbation that the majority of heritable gene expression variations is ''buffered'' and does not lead to the downstream effects on other genes or phenotypes. Phenotypic buffering of protein-coding polymorphisms has been already documented and it represents a general property of complex gene-regulatory networks (90–92). Most likely, common alleles are predominantly buffered by the robust properties of the system and then largely ''neutral'' for the rest of the molecules in the system. This scenario may have profound consequences for the design and interpretation of eQTL studies in relation to complex disease such as cancer. It could turn out that these complex diseases are not necessarily the result of common small-effect variants in a large number of genes but are rather caused by changes at a few crucial fragile points of the system (hotspots), which cause large, system-wide disturbances (93,94). The explanation of the rarity of eQTL master regulatory loci needs to be elucidated in future studies.

Conclusions

The complexity of CRC like other cancers has made understanding of underlying genetic factors difficult. The success of GWAS in identifying various genetic loci associated with the disease has opened new avenues from the perspective of understanding the genesis of the disease and its prevention and treatment. However, a gap in understanding the role of identified variants through GWAS remains to be fulfilled before any practical utility. Those identified variants being associated with expression regulation through *cis*- and *trans*-action remains a probability. Scarcity of information available about master regulators of expression and cancer risk association is mainly due to the difficulties in performing eQTL studies on cell lines other than the lymphoblastoid ones. eQTL studies may be a powerful tool for clarifying the role of candidate loci identified in GWAS. Master regulators can be more pertinent to provide information about the loci widely involved in the disease risk and in an extension also about the target genes and therefore, legitimate targets for future investigations.

Acknowledgements

Conflict of interest statement: None declared.

Funding

Grant Agency of the Czech Republic (CZ:GA CR:GA305/09/P194, CZ:GA CR:GAP304/10/1286).

References

1. Li, F. Y. and Lai, M. D. (2009) Colorectal cancer, one entity or three. *J. Zhejiang Univ. Sci. B.*, **10**, 219–229.
2. Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C. and Parkin, D. M. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer*, **127**, 2893–2917.
3. Hemminki, K., Lorenzo Bermejo, J. and Forsti, A. (2006) The balance between heritable and environmental aetiology of human disease. *Nat. Rev. Genet.*, **7**, 958–965.
4. Joshi, A. D., Corral, R., Siegmund, K. D. *et al.* (2009) Red meat and poultry intake, polymorphisms in the nucleotide excision repair and mismatch repair pathways and colorectal cancer risk. *Carcinogenesis*, **30**, 472–479.
5. Chan, A. T. and Giovannucci, E. L. (2010) Primary prevention of colorectal cancer. *Gastroenterology*, **138**, 2029–2043. e2010.
6. Ferlay, J., Parkin, D. M. and Steliarova-Foucher, E. (2010) Estimates of cancer incidence and mortality in Europe in 2008. *Eur. J. Cancer*, **46**, 765–781.
7. Fearon, E. R. and Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell*, **61**, 759–767.
8. Pritchard, C. C. and Grady, W. M. Colorectal cancer molecular biology moves into clinical practice. *Gut*, **60**, 116–129.
9. Prenen, H., Tejpar, S. and Van Cutsem, E. (2010) New strategies for treatment of KRAS mutant metastatic colorectal cancer. *Clin. Cancer Res.*, **16**, 2921–2926.
10. de la Chapelle, A. (2004) Genetic predisposition to colorectal cancer. *Nat. Rev. Cancer*, **4**, 769–780.
11. Aaltonen, L. A., Salovaara, R., Kristo, P. *et al.* (1998) Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N. Engl. J. Med.*, **338**, 1481–1487.
12. Samowitz, W. S., Curtin, K., Lin, H. H., Robertson, M. A., Schaffer, D., Nichols, M., Gruenthal, K., Leppert, M. F. and Slattery, M. L. (2001) The colon cancer burden of genetically defined hereditary nonpolyposis colon cancer. *Gastroenterology*, **121**, 830–838.
13. Cheah, P. Y. (2009) Recent advances in colorectal cancer genetics and diagnostics. *Crit. Rev. Oncol. Hematol.*, **69**, 45–55.
14. Tomlinson, I. P., Dunlop, M., Campbell, H. *et al.* (2010) COGENT (COlorectal cancer GENeTics): an international consortium to study the role of polymorphic variation on the risk of colorectal cancer. *Br. J. Cancer*, **102**, 447–454.
15. Hemminki, K., Li, X., Sundquist, J. and Sundquist, K. (2009) Cancer risks in Crohn disease patients. *Ann. Oncol.*, **20**, 574–580.
16. Ahmadi, A., Polyak, S. and Draganov, P. V. (2009) Colorectal cancer surveillance in inflammatory bowel disease: the search continues. *World J. Gastroenterol*, **15**, 61–66.
17. Bodmer, W. F. (2006) Cancer genetics: colorectal cancer as a model. *J. Hum. Genet.*, **51**, 391–396.
18. Speicher, M. R., Geigl, J. B. and Tomlinson, I. P. (2010) Effect of genome-wide association studies, direct-to-consumer genetic testing, and high-speed sequencing technologies on predictive genetic counselling for cancer risk. *Lancet Oncol.*, **11**, 890–898.
19. Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. and Hemminki, K. (2000) Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.*, **343**, 78–85.
20. Aaltonen, L., Johns, L., Jarvinen, H., Mecklin, J. P. and Houlston, R. (2007) Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. *Clin. Cancer Res.*, **13**, 356–361.
21. de Jong, M. M., Nolte, I. M., te Meerman, G. J., van der Graaf, W. T., de Vries, E. G., Sijmons, R. H., Hofstra, R. M. and Kleibeuker, J. H. (2002) Low-penetrance genes and their involvement in colorectal cancer susceptibility. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 1332–1352.
22. Chen, K., Jiang, Q. T. and He, H. Q. (2005) Relationship between metabolic enzyme polymorphism and colorectal cancer. *World J. Gastroenterol.*, **11**, 331–335.
23. Chao, C., Zhang, Z. F., Berthiller, J., Boffetta, P. and Hashibe, M. (2006) NAD(P)H:quinone oxidoreductase 1 (NQO1) Pro187Ser polymorphism and the risk of lung, bladder, and colorectal cancers: a meta-analysis. *Cancer Epidemiol. Biomarkers Prev.*, **15**, 979–987.
24. Hubner, R. A. and Houlston, R. S. (2007) MTHFR C677T and colorectal cancer risk: a meta-analysis of 25 populations. *Int. J. Cancer*, **120**, 1027–1035.

25. Huang, Y., Han, S., Li, Y., Mao, Y. and Xie, Y. (2007) Different roles of MTHFR C677T and A1298C polymorphisms in colorectal adenoma and colorectal cancer: a meta-analysis. *J. Hum. Genet.*, **52**, 73–85.
26. Tan, X. L., Nieters, A., Kropp, S., Hoffmeister, M., Brenner, H. and Chang-Claude, J. (2008) The association of cyclin D1 G870A and E-cadherin C-160A polymorphisms with the risk of colorectal cancer in a case control study and meta-analysis. *Int. J. Cancer*, **122**, 2573–2580.
27. Zhang, D., Chen, C., Fu, X., Gu, S., Mao, Y., Xie, Y., Huang, Y. and Li, Y. (2008) A meta-analysis of DNA repair gene XPC polymorphisms and cancer risk. *J. Hum. Genet.*, **53**, 18–33.
28. Broderick, P., Carvajal-Carmona, L., Pittman, A. M. *et al.* (2007) A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat. Genet.*, **39**, 1315–1317.
29. Haiman, C. A., Le Marchand, L., Yamamato, J., Stram, D. O., Sheng, X., Kolonel, L. N., Wu, A. H., Reich, D. and Henderson, B. E. (2007) A common genetic risk factor for colorectal and prostate cancer. *Nat. Genet.*, **39**, 954–956.
30. Tomlinson, I., Webb, E., Carvajal-Carmona, L. *et al.* (2007) A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.*, **39**, 984–988.
31. Zanke, B. W., Greenwood, C. M., Rangrej, J. *et al.* (2007) Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.*, **39**, 989–994.
32. Houlston, R. S., Webb, E., Broderick, P. *et al.* (2008) Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat. Genet.*, **40**, 1426–1435.
33. Jaeger, E., Webb, E., Howarth, K. *et al.* (2008) Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat. Genet.*, **40**, 26–28.
34. Pittman, A. M., Webb, E., Carvajal-Carmona, L. *et al.* (2008) Refinement of the basis and impact of common 11q23.1 variation to the risk of developing colorectal cancer. *Hum. Mol. Genet.*, **17**, 3720–3727.
35. Tenesa, A., Farrington, S. M., Prendergast, J. G. *et al.* (2008) Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat. Genet.*, **40**, 631–637.
36. Tomlinson, I. P., Webb, E., Carvajal-Carmona, L. *et al.* (2008) A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.*, **40**, 623–630.
37. Spain, S. L., Cazier, J. B., Houlston, R., Carvajal-Carmona, L. and Tomlinson, I. (2009) Colorectal cancer risk is not associated with increased levels of homozygosity in a population from the United Kingdom. *Cancer Res.*, **69**, 7422–7429.
38. Houlston, R. S., Cheadle, J., Dobbins, S. E. *et al.* (2010) Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat. Genet.*, **42**, 973–977.
39. Lascorz, J., Forsti, A., Chen, B. *et al.* (2010) Genome-wide association study for colorectal cancer identifies risk polymorphisms in German familial cases and implicates MAPK signalling pathways in disease susceptibility. *Carcinogenesis*, **31**, 1612–1619.
40. Cui, Y., Li, H., Wu, Q., Zhang, T., Kong, D., Song, T., Ru, T., Chen, P. and Li, Q. (2011) Treatment of colorectal cancer with unresectable synchronous liver-only metastases with combined therapeutic modalities. *J. Gastrointest. Surg.*, **15**, 285–293.
41. Cui, R., Okada, Y., Jang, S. G. *et al.* (2011) Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. *Gut*, **60**, 799–805.
42. Figueiredo, J. C., Lewinger, J. P., Song, C. *et al.* (2011) Genotype-environment interactions in microsatellite stable/microsatellite instability-low colorectal cancer: results from a genome-wide association study. *Cancer Epidemiol. Biomarkers Prev.*, **20**, 758–766.
43. Tomlinson, I. P., Carvajal-Carmona, L. G., Dobbins, S. E. *et al.* (2011) Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet.*, **7**, e1002105.
44. Tenesa, A. and Dunlop, M. G. (2009) New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nat. Rev. Genet.*, **10**, 353–358.
45. Tuupanen, S., Turunen, M., Lehtonen, R. *et al.* (2009) The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat. Genet.*, **41**, 885–890.
46. Pittman, A. M., Naranjo, S., Jalava, S. E. *et al.* (2010) Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a cis-acting regulator of EIF3H. *PLoS Genet.*, **6**, e1001126.
47. Carvajal-Carmona, L. G., Cazier, J. B., Jones, A. M. *et al.* (2011) Fine-mapping of colorectal cancer susceptibility loci at 8q23.3, 16q22.1 and 19q13.11: refinement of association signals and use of in silico analysis to suggest functional variation and unexpected candidate target genes. *Hum. Mol. Genet.*, **20**, 2879–2888.
48. Niittymaki, I., Tuupanen, S., Li, Y., Jarvinen, H., Mecklin, J. P., Tomlinson, I. P., Houlston, R. S., Karhu, A. and Aaltonen, L. A. (2011) Systematic search for enhancer elements and somatic allelic imbalance at seven low-penetrance colorectal cancer predisposition loci. *BMC Med. Genet.*, **12**, 23.
49. Fransen, K., Visschedijk, M. C., van Sommeren, S. *et al.* (2010) Analysis of SNPs with an effect on gene expression identifies UBE2L3 and BCL3 as potential new risk genes for Crohn's disease. *Hum. Mol. Genet.*, **19**, 3482–3488.
50. Murphy, A., Chu, J. H., Xu, M. *et al.* (2010) Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood CD4+ lymphocytes. *Hum. Mol. Genet.*, **19**, 4745–4757.
51. Glinskii, A. B., Ma, J., Ma, S., Grant, D., Lim, C. U., Sell, S. and Glinsky, G. V. (2009) Identification of intergenic trans-regulatory RNAs containing a disease-linked SNP sequence and targeting cell cycle progression/differentiation pathways in multiple common human disorders. *Cell Cycle*, **8**, 3925–3942.
52. Nica, A. C., Montgomery, S. B., Dimas, A. S., Stranger, B. E., Beazley, C., Barroso, I. and Dermitzakis, E. T. (2010) Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.*, **6**, e1000895.
53. Glinskii, V. G. and Glinsky, G. V. (2008) Emerging genomic technologies and the concept of personalized medicine. *Cell Cycle*, **7**, 2278–2285.
54. Cheung, V. G., Spielman, R. S., Ewens, K. G., Weber, T. M., Morley, M. and Burdick, J. T. (2005) Mapping determinants of human gene expression by regional and genome-wide association. *Nature*, **437**, 1365–1369.
55. Dixon, A. L., Liang, L., Moffatt, M. F. *et al.* (2007) A genome-wide association study of global gene expression. *Nat. Genet.*, **39**, 1202–1207.
56. Dermitzakis, E. T. (2008) From gene expression to disease risk. *Nat. Genet.*, **40**, 492–493.
57. Cheung, V. G., Nayak, R. R., Wang, I. X., Elwyn, S., Cousins, S. M., Morley, M. and Spielman, R. S. (2010) Polymorphic cis- and trans-regulation of human gene expression. *PLoS Biol*, **6**, e1000480.
58. Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S. and Cheung, V. G. (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature*, **430**, 743–747.
59. King, M. C. and Wilson, A. C. (1975) Evolution at two levels in humans and chimpanzees. *Science*, **188**, 107–116.
60. Deutsch, S., Lyle, R., Dermitzakis, E. T. *et al.* (2005) Gene expression variation and expression quantitative trait mapping of human chromosome 21 genes. *Hum. Mol. Genet.*, **14**, 3741–3749.
61. Cheung, V. G. and Spielman, R. S. (2009) Genetics of human gene expression: mapping DNA variants that influence gene expression. *Nat. Rev. Genet.*, **10**, 595–604.
62. Jansen, R. C. and Nap, J. P. (2001) Genetical genomics: the added value from segregation. *Trends Genet.*, **17**, 388–391.
63. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. and Lathrop, M. (2009) Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.*, **10**, 184–194.
64. Mills, R. E., Walter, K., Stewart, C. *et al.* (2011) Mapping copy number variation by population-scale genome sequencing. *Nature*, **470**, 59–65.
65. Zeller, T., Wild, P., Szymczak, S. *et al.* (2010) Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS One*, **5**, e10693.
66. Li, Y., Sheu, C. C., Ye, Y. *et al.* (2010) Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. *Lancet Oncol.*, **11**, 321–330.
67. Choy, E., Yelensky, R., Bonakdar, S. *et al.* (2008) Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. *PLoS Genet.*, **4**, e1000287.
68. Williams, R. B., Chan, E. K., Cowley, M. J. and Little, P. F. (2007) The influence of genetic variation on gene expression. *Genome Res.*, **17**, 1707–1716.
69. Rockman, M. V. and Kruglyak, L. (2006) Genetics of global gene expression. *Nat. Rev. Genet.*, **7**, 862–872.
70. Buil, A., Perera-Lluna, A., Souto, R., Peralta, J. M., Almasy, L., Vallverdu, M., Caminal, P. and Soria, J. M. (2007) Searching for master regulators of transcription in a human gene expression data set. *BMC Proc.*, **1** (Suppl. 1), S81.
71. Chesler, E. J., Lu, L., Shou, S. *et al.* (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat. Genet.*, **37**, 233–242.

72. Brem, R. B., Yvert, G., Clinton, R. and Kruglyak, L. (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science*, **296**, 752–755.
73. Schadt, E. E., Monks, S. A., Drake, T. A. *et al*. (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature*, **422**, 297–302.
74. Yvert, G., Brem, R. B., Whittle, J., Akey, J. M., Foss, E., Smith, E. N., Mackelprang, R. and Kruglyak, L. (2003) Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. *Nat. Genet*., **35**, 57–64.
75. Bystrykh, L., Weersing, E., Dontje, B. *et al*. (2005) Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nat. Genet*., **37**, 225–232.
76. Hubner, N., Wallace, C. A., Zimdahl, H. *et al*. (2005) Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat. Genet*., **37**, 243–253.
77. Breitling, R., Li, Y., Tesson, B. M. *et al*. (2008) Genetical genomics: spotlight on QTL hotspots. *PLoS Genet*., **4**, e1000232.
78. Schadt, E. E., Molony, C., Chudin, E. *et al*. (2008) Mapping the genetic architecture of gene expression in human liver. *PLoS Biol*., **6**, e107.
79. Monks, S. A., Leonardson, A., Zhu, H., Cundiff, P., Pietrusiak, P., Edwards, S., Phillips, J. W., Sachs, A. and Schadt, E. E. (2004) Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet*., **75**, 1094–1105.
80. Stranger, B. E., Forrest, M. S., Clark, A. G. *et al*. (2005) Genome-wide associations of gene expression variation in humans. *PLoS Genet*., **1**, e78.
81. Goring, H. H., Curran, J. E., Johnson, M. P. *et al*. (2007) Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat. Genet*., **39**, 1208–1216.
82. Emilsson, V., Thorleifsson, G., Zhang, B. *et al*. (2008) Genetics of gene expression and its effect on disease. *Nature*, **452**, 423–428.
83. de Koning, D. J. and Haley, C. S. (2005) Genetical genomics in humans and model organisms. *Trends Genet*., **21**, 377–381.
84. Benfey, P. N. and Mitchell-Olds, T. (2008) From genotype to phenotype: systems biology meets natural variation. *Science*, **320**, 495–497.
85. Wu, C., Delano, D. L., Mitro, N. *et al*. (2008) Gene set enrichment in eQTL data identifies novel annotations and pathway regulators. *PLoS Genet*., **4**, e1000070.
86. Stylianou, I. M., Affourtit, J. P., Shockley, K. R., Wilpan, R. Y., Abdi, F. A., Bhardwaj, S., Rollins, J., Churchill, G. A. and Paigen, B. (2008) Applying gene expression, proteomics and single-nucleotide polymorphism analysis for complex trait gene identification. *Genetics*, **178**, 1795–1805.
87. Zhu, J., Zhang, B., Smith, E. N., Drees, B., Brem, R. B., Kruglyak, L., Bumgarner, R. E. and Schadt, E. E. (2008) Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat. Genet*., **40**, 854–861.
88. Gibson, G. and Dworkin, I. (2004) Uncovering cryptic genetic variation. *Nat. Rev. Genet*., **5**, 681–690.
89. Le Rouzic, A. and Carlborg, O. (2008) Evolutionary potential of hidden genetic variation. *Trends Ecol. Evol*., **23**, 33–37.
90. Rutherford, S. L. and Lindquist, S. (1998) Hsp90 as a capacitor for morphological evolution. *Nature*, **396**, 336–342.
91. Queitsch, C., Sangster, T. A. and Lindquist, S. (2002) Hsp90 as a capacitor of phenotypic variation. *Nature*, **417**, 618–624.
92. Bergman, A. and Siegal, M. L. (2003) Evolutionary capacitance as a general feature of complex gene networks. *Nature*, **424**, 549–552.
93. Iyengar, S. K. and Elston, R. C. (2007) The genetic basis of complex traits: rare variants or ''common gene, common disease''? *Methods Mol. Biol*., **376**, 71–84.
94. Bodmer, W. and Bonilla, C. (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet*., **40**, 695–701.