ELECTRONIC LETTER

Single nucleotide polymorphism (SNP) analysis of mouse pulmonary adenoma susceptibility loci 1–4 for identification of candidate genes

W J Lemon, C H Swinton, M Wang, N Berbari, Y Wang, M You

....

L ung cancer is the leading cause of mortality from cancer in both men and women in developed countries. There is evidence that, although incidence is almost always associated with environmental factors such as smoking or occupational exposure, susceptibility has a genetic component with early onset lung cancer following Mendelian inheritance.[1 2] Moreover, susceptibility is largely intrinsic to the lung itself as shown by the classical experiments involving lung explants from sensitive and resistant mice.[3] Identification of the genes predisposing to cancer could yield targets for treatment or chemoprevention. In humans, the wide variety of carcinogens and varying degrees of exposure make identifying the predisposing genes difficult, but in a mouse model, such confounding variables can be controlled.

Previously, classical genetic studies involving cross breeding of mouse strains with differing susceptibilities have identified

chromosomal areas associated with predisposition to developing spontaneous and chemically induced lung adenomas.[4 5] The process involves use of many evenly spaced polymorphic DNA markers to create landmarks across each chromosome. In the process called quantitative trait loci (QTL) mapping, attempts are made to find significant correlations between marker alleles and the phenotypic variation, or disease state.[6] Inbred mouse strains vary in their susceptibility to cancer and two extreme strains are A/J (susceptible) and C57BL/6J (B6, resistant).[4 5] Four QTLs identified as pulmonary adenoma susceptibility (Pas) loci 1–4 have been mapped, respectively, to mouse chromosomes 6, 17, 19, and 9. These regions are illustrated in fig 1.

Much genetic variation between people is a result of random mutation at specific nucleotide positions. These single nucleotide polymorphisms (SNPs, pronounced snips) have the potential to produce profound effects on gene expression and consequently phenotype, as shown in fig 2. For example, SNPs in the 3′ UTR region can alter the stability of the mRNA, by changing binding sites or secondary structure, thus making it more or less likely to be degraded. A SNP in the 5′ region can change promoter binding sites and thereby modify the affinity for a transcription factor. Nonsense mutations are the most severe of the SNPs found in the coding region. Nonsense SNPs are ones that introduce a premature stop codon resulting in a truncated polypeptide, often resulting in loss of function. Missense mutations result in an amino acid change that can be important if the properties of the new amino acid (charge, polarity, etc) are different from the one it replaced. As susceptibility to lung cancer has a strong genetic component, detailed SNP analysis of polymorphisms between A/J and B6 (A-B SNPs) may facilitate the identification of candidate susceptibility genes. The genomes of both strains have been sequenced and assembled, and thus direct comparison is now possible.

Celera's mouse reference SNP database (1.0), the first such resource to become available, includes 2 566 706 SNPs and is based on the first *Mus musculus* assembly release which includes the publicly produced sequence of C57BL/6J and the Celera produced sequences of A/J, DBA/2J (DBA), 129X1/SvJ, and 129S1/SvImJ. The database was scanned for A-B SNPs in the Pas1–4 QTLs. The resulting list of SNPs was analysed to identify the most likely candidate genes. This SNP analysis complements previous gene expression analysis and, as in that work, builds on classic genetic results and integrates newly available genomic data.[7]

Key points

- Past studies have mapped four susceptibility loci (Pas1–4) for pulmonary adenoma in which A/J and C57BL/6J (B6) mice have different alleles that affect incidence and multiplicity of tumours. With the release of a genome wide SNPs database, it has become feasible to analyse these genetically determined QTLs for genes polymorphic in these strains.
- Celera's discovery database (CDS 3.6, SNP 1.0) was scanned for SNPs in the Pas1–4 QTLs. SNPs were first screened according to the following criteria: (1) A/J and B6 were polymorphic for the SNP; (2) SNPs appeared in the coding region, the 5′ regulatory region, or the 3′ untranslated region; (3) SNPs appeared in known genes; (4) B6 and DBA/2J, phenotypically similar to B6, shared the same allele.
- Genes for which associations or other plausible links with cancer have been published were deemed as final candidates. All 11 selected SNPs within candidate genes were verified by polymerase chain reaction (PCR) sequencing. We have also attempted to verify a series of differentially expressed candidate susceptibility genes to lung tumours in our previous microarray analysis with semiquantitative reverse transcriptase PCR (RT-PCR).
- Differential expression of six of seven candidate genes were confirmed. Candidate Pas genes are Pas1, *receptor type protein tyrosine phosphatase* and *basic helix-loop-helix B3*; Pas2, *Notch4*, *CREBL1*; Pas3, *Minpp1* and *FoxD4*; and Pas4, *TGF β receptor II*. Identification of the genes predisposing to mouse lung cancer could have considerable implications for diagnosis, treatment, or chemoprevention of lung cancer in humans.

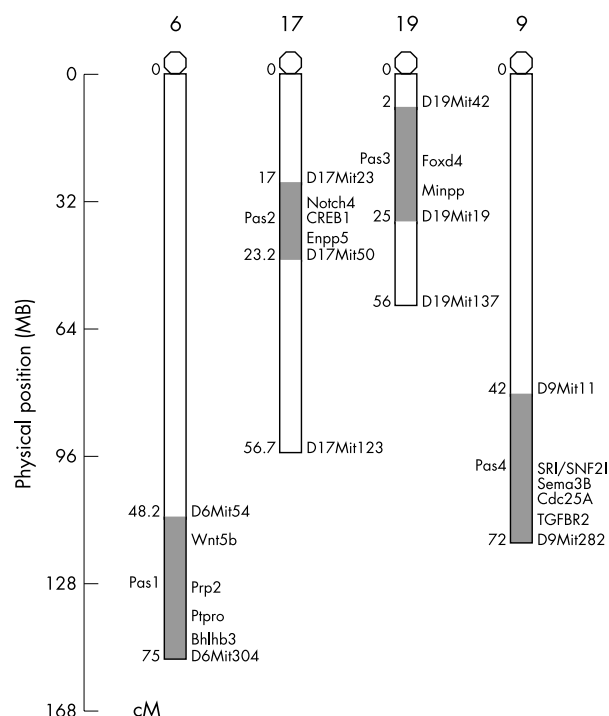...

Figure 1 Mouse pulmoary adenoma susceptibility loci (Pas) as mapped in crosses of strains A/J and C57BL/6. Diagrams show chromosomes drawn to a scale of physical length according to Celera CDS 3.6. Pas1 appears at the distal end of chromosome 6 with genetic position of the flanking markers at the left and the marker names and key genes at the right. Pas2–4, respectively on chromosomes 17, 19, and 9, are similarly labelled.

Figure 2 Hypothetical examples of biologically effective single nucleotide polymorphisms (SNPs). Disease producing SNPs in the untranslated regions may influence mRNA stability in the case of 3′ UTR (top panel) or promoter activity in the 5′ UTR (second panel). Among the coding SNPs, missense SNPs (third panel) could cause an amino acid change that, for example, removes a disulphide bridge. Nonsense SNPs, by definition, introduce truncations that can diminish or obliterate protein function.

MATERIALS AND METHODS

Approach

Markers flanking each QTL were located in the Celera (Rockville, MD) mouse genome database (www.celera.com, CDS 3.6 release). These markers were used to identify the reference DNA positions within the assembled genome for our database query. The Celera mouse SNP reference database (v1.0) was queried for SNPs within each QTL then filtered to keep only A-B SNPs.

Data filtering

Additional filtering of the A-B SNPs was performed to increase confidence in positive results. Often with screening one seeks to reduce the false negative rate and thereby be inclusive, but owing to the tremendous number of SNPs (~20 000 in Pas1 alone), we sought to minimise the false positive rate. To accomplish this, A-B SNPs that met the following criteria were selected for further analysis: (1) it appeared in the coding region, the 5′ regulatory region,

or the 3′ untranslated region; (2) it appeared in a known gene; (3) the DBA (another resistant strain) allele was the same as the B6 allele. Intronic, intergenic, and silent SNPs were excluded. This was done for two reasons: (1) the function of intronic and intergenic sequence is a matter of research beyond the scope of this study; (2) the number of these SNPs made their analysis currently untenable. For the resulting genes containing A-B SNPs, further analysis was performed to sequence verify the SNPs with PCR sequencing analysis, and to identify those genes or gene families in which association with lung cancer has been previously established.

Table 1 Summary of SNPs found in Celera Mouse SNP Reference database (v 1.0). Total refers to the number of all SNPs (any type, any strain-strain polymorphism) in the region. Filtered refers to the number of SNPs meeting the criteria described in the text. "No of genes" refers to the number of genes associated with the filtered SNPs. Known genes are any genes having specific functional annotation in the Celera CDS 3.6 database

Locus	Total	Filtered	No of genes	Known Genes	No of genes
Pas1	19 092	99	65	16	15
Pas2	19 639	69	37	25	14
Pas3	26 916	96	55	16	11
Pas4	21 505	147	88	24	14

Table 2 A–B SNPs in known Pas1 genes. SNPs appear in order of position on the chromosome. SNP ID is Celera's unique ID number for the SNP. GenBank refers to the transcript sequence associated with the gene containing the SNP. Gene name is derived from Celera's annotation wherever possible. Type refers to the gene structure containing the SNP and, for coding SNPs, it refers to the effect on amino acid sequence predicted by the most authoritative reading frame(s) for the transcript. Chr refers to the base pair position of the SNP on the chromosome according to the Celera CDS 3.6 assembly. mRNA refers to its nucleotide position within the corresponding Celera transcript. Allele/codon describes the nucleotide change, and, when it occurs within the translated region, it is described in terms of the codon triplet. C57BL/6J alleles always appear on the left, A/J alleles on the right. DBA refers to concordance of DBA/2J allele with B6 allele. RT is checked when the SNP was confirmed by RT-PCR sequencing. Array refers to whether the gene was found differentially regulated in a previous microarray study (Lemon WJ et al. J Med Genet 2002;**39**:644–55). Pb is checked when published reports have associated the gene with cancer. Description is derived from either Celera annotation or GenBank annotation, whichever the authors deemed more clear, with deference to Celera.

SNP ID	GenBank	Gene name	Type	Relative position		Codon	AA	DBA	SNP RT	Pb	Description
				Chr	mRNA						
mCV22281390	BM489316	Arpc4	Missense	112,205,003	48	GGC GTC	Gly Val				Actin related protein 2/3 complex, subunit 4
mCV22584830	NM_009525	Wnt5b	3′ UTR	118,266,570	1158	A/G				Y	Wingless related MMTV integration site 5B
mCV25250691	NM_008646	Mug2	Nonsense	120,923,657	4417	TGA CGA	Stop Arg	Y			Murinoglobulin 2
mCV25250751	M65237	Mug1	5′ UTR, missense	120,988,713	788/808	GTA GCA	Ala Val				Murinoglobulin (MUG3)
mCV24835024	NM_009645	Aicda	3′ UTR	121,357,737	1703	G/A					Activation induced cytidine deaminase
mCV24136983	NM_027218	1810046I24Rik	Missense	121,870,184	618	AAA AAC	Asn Lys	Y			C type lectin
mCV22899606	NM_010736	Ltbr	3′ UTR	123,978,076	1445	G/A					Lymphotoxin B receptor
mCV25230505	NM_007376	A2m	Missense	127,183,453	1670	ATT GTT	Ile Val	Y			Alpha-2-macroglobulin
mCV22524446	NM_031499	Prp2	Missense	130,318,725	5252	CAA GAA	Glu Gln	Y			Proline rich protein 2
mCV24262005	NM_008462	Klra2	3′ UTR	133,237,819	1237	C/T		Y	Y		Killer cell lectin-like receptor, subfamily A, member 2
mCV23359715	NM_026371	5830457J20Rik	3′ UTR	134,662,880	1095	T/A		Y	Y		
mCV23359716			3′ UTR	134,663,400	1615	A/G		Y			
mCV24045161	NM_025589	2410038A03Rik	Missense	136,486,200	201	GAT GGT	Asp Gly	Y			
mCV23381606	NM_011216	Prpro	3′ UTR	137,526,924	1929	-/T		Y	Y		Protein tyrosine phosphatase, receptor type, O
mCV24099286	NM_024469	Bhlhb3	Missense	145,935,561	812	TCG CCG	Ser Pro	Y	Y	Y	basic helix-loop-helix domain containing, class B3

Table 3 A–B SNPs in Pas2 known genes. Columns as described for table 2.

SNP ID	GenBank	Gene name	Type	Relative position		Allele/codon	AA	DBA	SNP RT	Description
				Chr	mRNA					
mCV24100604	NM_020581	Angptl4	Ins/del	33,532,739	1318	G/-		Y		Angiopoietin-like 4
mCV24103348	NM_030697	0610013D04Rik	Missense	33,577,259	1809	TCT CCT	Ser Pro	Y		
mCV24105091	NM_016761	Kif1c	3′ UTR	33,635,219	1545	G/C		Y		Kinesin family member C1
mCV24107583	NM_020603	Bing4	Missense	33,693,845	118	ACC GCC	Thr Ala	Y	Y*	Sequence of the mouse major histocompatibility locus class II region
mCV24297946	NM_010929	Notch4	Missense	34,676,126	3014	GAA GAC	Glu Asp	Y	Y	Common integration site for the mouse mammary tumour virus
mCV24296263	NM_017406	Crebl1	Missense	34,745,777	662	AAC AGC	Asn Ser	Y	Y	cAMP responsive element binding protein-like 1, ubiquitous transcription factor
mCV24296262			Missense	34,745,896	697	TCT GCT	Ser Ala	Y		
mCV24295885	NM_031176	Tnxb	Missense	34,767,693	1838	CAC CGC	Asn Arg	Y		Tenascin XB
mCV22965429	NM_020027	Bat2	Missense	35,246,204	6470	CAC CGC	His Arg	Y		HLA-B associated transcript 2 (aka HLA-G2)
mCV22965384			Acceptor splice site	35,246,540	6243	T/C				
mCV23273805	NM_008199	H2-Bl	Nonsense	36,632,166	1023	TAG CAG	Asn Gln			Histocompatibility 2, blastocyst - expressed at the blastocyst stage & placenta
mCV23273810			Missense	36,634,076	827	AAG AAT	Lys Asn			
mCV23273811			Missense	36,634,087	816	GAG CAG	Asn Gln			
mCV23273815			Missense	36,634,119	784	GCC GGC	Gly Ala			
mCV23273817			Missense	36,635,104	180	GGT AGT	Gly Ser			
mCV24433115			3′ UTR	36,698,819	387	T/C		Y		
mCV24433812	NM_025520	2310034K10Rik	3′ UTR	36,732,547	1929	C/T				
mCV24436558	NM_024468	Rnf23	Missense	37,495,966	2476	ACA GCA	Thr Ala	Y		Tripartite motif protein 39
mCV23715424	NM_019439	Gabbr1	Missense	41,344,966	396	GTA ATA	Val Ile	Y		Gamma-aminobutyric acid (GABA-B) receptor, 1, note: putative CpG island
mCV23715425	BM490573	2610528M18Rik	Missense	41,344,981	381	GTT ATT	Val Ile	Y		
mCV25200870	NM_032003	Enpp5	Missense	44,507,091	1530	CAT AAT	His Asn	Y	N	Ectonucleotide pyrophosphatase/phosphodiesterase 5
mCV25200871			3′ UTR	44,507,302	1741	G/A		Y		
mCV25200881			3′ UTR	44,507,604	2043	A/G		Y		
mCV25200882			3′ UTR	44,507,769	2208	C/-		Y		

*A/J allele not reported in Celera SNP Reference Database 1.0.

Table 4 A-B SNPs in Pas3 known genes. Columns as described for table 2

| | | | | Relative position | | | | | | | |
| | | | | Chr | mRNA | | | | | | |
SNP ID	GenBank ID	Gene name	Type	Chr	mRNA	Allele/codon	AA	DBA	SNP RT	Pb	Description
mCV23294255	NM_007650	Cd5	3′ UTR	8,418,081	1811	T/C			Y		CD5 antigen
mCV23294266			3′ UTR	8,418,184	1708	T/C			Y		
mCV23294598			Missense	8,426,359	100	GTC ATC	Val Ile				
mCV23010981	NM_025658	2210417J23Rik	Missense	9,221,967	437	GTT GCT	Val Ala				
mCV24416503	NM_008139	Gnaq	3′ UTR	14,037,739	1241	T/G					Guanine nucleotide binding protein, alpha q polypeptide
mCV23044840	NM_010730	Anxa1	Missense	18,100,327	946	GAA GGA	Glu Gly				Annexin A1
mCV24486668	NM_008022	Foxd4	Missense	22,674,812	1698	CAG CGG	Gln Arg		Y		Forkhead box D4
mCV24486689			Missense	22,675,284	1226	GGG CGG	Gly Arg	Y	Y	Y	
mCV22823123	NM_028785	1200017A24Rik	3′ UTR	22,988,476	2509	A/G					
mCV23196871	NM_013703	Vldlr	Missense	25,022,920	1118	GGG AGG	Gly Arg				Very low density lipoprotein receptor
mCV23093281	NM_031249	Cstf2t-pending	3′ UTR	28,869,655	2056	G/C			Y		Cleavage stimulation factor, 3′ pre-RNA subunit 2
mCV23094584	NM_018830	Asah2	5′ UTR	29,849,129	306	T/G			Y		N-acylsphingosine amidohydrolase 2
mCV25400117	BB002878	Minpp1	Nonsense	30,268,848	564	TGA TGG	Stop Trp		Y		Multiple inositol polyphosphate histidine phosphatase 1
mCV25441217	NM_026487	4921525H23Rik	3′ UTR	30,455,931	1840	C/T			Y		
mCV25441216			3′ UTR	30,455,946	1825	G/A			Y		
mCV25441200			3′ UTR	30,456,513	1258	A/G					

Table 5 A-B SNPs in Pas4 known genes. Columns are as described for table 2

| | | | | Relative position | | | | | | | |
| | | | | Chr | mRNA | | | | | | |
SNP ID	GenBank	Gene Name	Type	Chr	mRNA	Alelle/codon	AA	DBA	SNP RT	Pb	Description
mCV24278514	NM_013669	Snap91	UTR 5	80,912,581	855	G/A					Synaptosomal associated protein, 91
mCV24278524			UTR 5	80,912,656	780	A/G					
mCV23421071	NM_025360	1200002G13Rik	UTR 3	86,317,238	768	T/C					
mCV25420569	NM_011961	Plod2	UTR 3	89,191,569	2,231	C/T					Procollagen lysine, 2-oxoglutarate 5-dioxygenase 2
mCV23564270	NM_024291	Ky	UTR 3	99,085,849	4114	A/G					Kyphoscoliosis
mCV23564271			UTR 3	99,085,879	4144	G/A					
mCV23586226	NM_025371	1110014J22Rik	Missense	103,062,990	533	GTT GCT	Val Ala	Y			Aminoacylase 1
mCV23592429	NM_030730	Srisnf2l-pending	UTR 3	103,313,283	4819	G/A		Y	Y		Steroid receptor-interacting SNF2 domain protein
mCV23592440			Missense	103,314,355	3747	GTC ATC	Val Ile	Y	Y		
mCV24367303	NM_010489	Hyal2	UTR 3	104,181,371	1730	T/C		Y		Y	Hyaluronidase 2
mCV22576892	NM_025903	Ifrd2	Missense	104,196,642	152	TCC CCC	Ser Pro				Interferon related developmental regulator 2
mCV22576749			Missense	104,198,775	329	GCC ACC	Ala Thr				
mCV22576540	BM486632	Sema3b	UTR 3	104,207,874	2139	T/A			Y		
mCV25006051	NM_011634	Traip	Missense	104,559,428	46	GCC ACC	Ala Thr				TRAF interacting protein
mCV25024853	NM_007658	Cdc25a+I154	Missense	106,607,499	384	CAC CAA	His Gln		Y		Cell division cycle 25 homologue A
mCV23789456	NM_017466	Cmkbr1l2	UTR 5	107,780,575	190	C/A					Chemokine (C-C) receptor 1-like 2
mCV23789449			UTR 5	107,780,576	189	C/T					
mCV23789448			UTR 5	107,780,703	62	G/A					
mCV23789447			UTR 5	107,780,748	17	A/C					
mCV23303603	NM_009371	Tgfbr2	UTR 3	113,551,165	2312	T/G		Y*	N*	Y	Transforming growth factor, beta receptor II
mCV23303602			UTR 3	113,551,334	2143	C/T		Y*	N*		
mCV22520381	NM_025974	3100001N19Rik	Missense	118,043,888	568	GTT GCT	Val Ala				
mCV22520379			UTR 3	118,044,104	784	A/G					
mCV22520373			UTR 3	118,044,148	828	T/C					

*See table 3.

SNP annotation

The SNP IDs were used to link SNPs with the genes in which they occur and Celera's mechanism to link its genes with public sequences was used for reporting public accession numbers and descriptions from GenBank (RefSeq when possible).

Sequence verification of SNPs

A-B SNPs were sequence verified with RT-PCR sequencing by the sequencing core facility in the Division of Human Cancer Genetics, Ohio State University. Briefly, the transcript sequence for the candidate genes were downloaded from the Celera database and PCR primers flanking the SNPs were designed such that 200–400 bp fragments were produced. RT-PCR, as described later, was performed to amplify total RNA harvested from lungs of A/J and C57BL/6J mice. These amplified fragments were submitted to the sequencing core facility where they were processed according to standard procedures.

Confirmation of differential gene expression by RT-PCR

Expression of several candidate genes found here by SNP analysis, or previously by array analysis,[7] were verified by RT-PCR. RNA was harvested from normal lungs of A/J, BALB/cJ, and C57BL/6J mice, reverse transcribed and amplified by PCR. A mixture of 4 μg of total RNA, oligo (dT) primer (Invitrogen, Grand Island, NY, USA), and double distilled water was incubated at 70°C for 10 minutes. This mixture was then added to the reverse transcription reaction mixture containing the following: forward reaction buffer, DTT (Invitrogen, Grand Island, NY, USA), RNasin (Promega, Madison, WI, USA), dNTPs, and M-MLV reverse transcriptase (Invitrogen, Grand Island, NY, USA). This final mixture was incubated at 37°C for one hour. Next, a final incubation at 95°C for two minutes was performed to remove RNA.

For real time PCR, a LightCycler-DNA Master SYBR Green I kit (Roche, Indianapolis, IN, USA) and an iCycler iQ (Biorad, Rodeo, CA, USA) thermocycler were used. Each reaction on a 96 well microtitre plate contained 2 μl of cDNA template and 2 μl of master mix. The kit master mix consisted of *Taq* DNA polymerase, reaction buffer, dNTP mix, SYBR Green I dye, and 10 mmol/l MgCl$_2$. The final reaction had a 1 μmol/l concentration of forward and reverse primers and 2 mmol/l MgCl$_2$ concentration, and a final volume of 20 μl. The PCR variables were as follows: a heating temperature of 95°C, an annealing temperature of 55°C, and an elongation temperature of 72°C were used for 34 cycles.

The genes examined were *SmarcD3*, *ATFa*, *Cdk5*, *PDCD4*, *Cdc25A*, and *Mbd2* together with β actin as a control. Base 2 logarithm of fold changes were computed, for gene x between strains A and B, as $\log_2(x_A)-\log_2(\beta\ actin_A)+\log_2(\beta\ actin_B)-\log_2(x_B)$.

RESULTS

Our approach to identifying candidate susceptibility genes for lung tumours with SNPs complements our earlier work with gene expression analysis.[7] As in the previous study, this work builds on existing genetic mapping of QTLs, availability of the mouse genome sequence, and SNP data. The genome sequence, combined with previously identified QTLs, allows focus on SNP analysis within the regions known to modulate susceptibility and resistance.

Fig 1 illustrates the physical positions of the Pas QTLs, highlights the flanking markers and shows the corresponding genetic positions. As expected, proximal regions span less genetic distance than equal length distal regions, owing to the higher distal recombination rate. Table 1 shows the number of SNPs found in each QTL and the number of genes containing SNPs in each QTL for each step in the filtering process. For example, 19 092 SNPs were found in Pas1. After removing the silent, intronic, and intergenic SNPs and those not A-B polymorphic, the number was reduced to 99 A-B SNPs associ-

ated with 65 genes. After the final step of isolating only the A-B SNPs in known genes, 16 SNPs in 15 genes remained.

Tables 2–5 enumerate, for each Pas QTL, the A-B SNPs found in known genes. Of the 15 A-B genes containing SNPs in Pas1, two stand out as strong candidates: *Ptpro* and *Bhlhb3*. In both genes, the B6 allele is shared with DBA and the sequence of the SNP for *Ptpro* has been verified. *Ptpro* is a receptor type O protein, tyrosine phosphatase. It has been suggested that other members of this family are tumour suppressor genes, including the recent suggestion that *Ptprj* (mouse chromosome 2) is the susceptibility gene to colon cancer 1 (*Scc1*).[8] *Bhlhb3* contains a basic helix-loop-helix domain and is expressed in late embryogenesis and may be a target of *Notch* ligands.[9] [10]

Pas2

In this locus, often noted for the preponderance of HLA genes, two non-HLA genes stand out. *Notch4*, as a known oncogene,[11] and *CrebL1*, as a member of the cyclic AMP response element binding proteins, could contribute to metastatic potential of melanoma cells.[12] [13] The A/J allele for the *Notch4* SNP was not reported in the Celera database, but sequencing showed it to vary between A/J and B6. Both seem to have a functional relation to tumorigenesis and the B6 allele is shared with DBA.

Pas3

Two genes were found to be candidates, *Minpp1* and *FoxD4*. The sequences of SNPs for both genes were verified. *Forkhead box D4* is the best candidate, having one B6 allele shared with DBA and being a member of a family of transcription factors that regulate the cell cycle, and it may play a part in tumorigenesis.[14]

Pas4

TGFβRII has historically been the primary candidate for Pas4. Both B6 alleles are shared with DBA, but the A/J allele is not reported in the Celera database. As an SNP was found in Celera between B6 and 129X1, we checked the region for a SNP between A/J and B6, but found none. It has been shown that loss of the type II receptor is often associated with a loss of TGF-β induced cell cycle repression.[15] Reduction of type II receptor mRNA is strongly associated with lung adenocarcinoma in mice and humans.[15] [16] However, RT-PCR sequencing showed no SNP between A/J and B6.
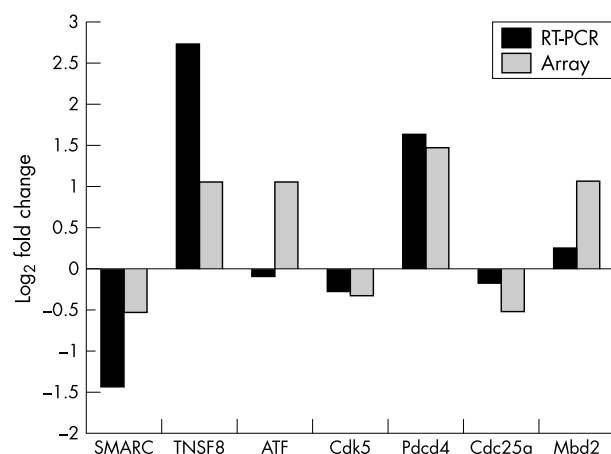
Figure 3 Comparison of microarray data with RT-PCR for selected genes. Fold change of gene expression between A/J and C57BL/6J for five genes and between A/J and BALB/cJ for one gene (*Mbd2*) are compared on a log$_2$ basis. Each unit on the y axis represents a twofold change. Negative values represent higher expression in B6.

RT-PCR verification of microarray data

Fig 3 shows RT-PCR follow up on selected genes found to be candidates in our previous microarray study.[7] In that study, RNAs from normal mouse tissue from these strains were evaluated for differential expression across strains. Genes within the QTLs that are differentially expressed across strains were deemed as candidates on the basis that expression differences could predispose to tumorigenesis. However, the microarray data were not confirmed by conventional RT-PCR or northern blotting techniques. In this study, we selected seven of the differentially expressed candidate genes found in that study for further confirmation. Six of the seven genes examined by RT-PCR showed differential expression similar to that found with the array, as shown in fig 3. One gene, *ATFa*, did not show the same pattern with RT-PCR and array. Note that this gene had the lowest level of expression of the seven genes (Li-Wong full model estimates of gene expression: B6=27, A/J=58). Fig 3 shows the correspondence of differential expression measured in the array data with that found by RT-PCR and correspondence between the two assay systems is consistent for six of seven candidate genes assayed.

DISCUSSION

Following the commonly held and straightforward view that genetic determinants of susceptibility may arise from the simplest form of genetic variation, the SNP, we analysed SNPs in four susceptibility loci to pulmonary adenoma. Five strains have been sequenced and assembled to date: A/J (susceptible, s), DBA/2J (resistant, r), C57BL/6J (r), and two substrains of 129, namely 129X1/SvImJ (intermediate, i) and 129S1/SvJ (i).[5 17] Substrain 129X1/SvJ was contaminated in 1978 and, as a result, contains as much as 25% non-129 DNA. Owing to consequent large scale heterozygosity, it is considered by some as not an inbred strain.[18–20] Manenti *et al*[17] reported 129X1/SvJ to be of intermediate susceptibility which leaves the expectation for Pas genotypes ambiguous. Similarly strain 129S1, the "steel" strain, is an inbred strain, but being of intermediate susceptibility, the expectation for Pas genotypes is also ambiguous. Consequently, our criteria required that DBA and B6 share alleles and that A/J be different, but the 129 alleles were ignored. Analysis was limited to SNPs likely to be detectable in mRNA, namely, coding SNPs, splice site SNPs, and UTR SNPs. As *K-ras*, which has been a strong Pas1 candidate gene for some time, has polymorphisms between A/J and B6 in intron 2,[5] we would have liked to include intronic SNPs. However, so many intronic SNPs were identified that it was deemed at this point untenable to sort out the useful ones.

Pas1 has been shown to be the major locus, accounting for 45% of the phenotypic variance.[4] Mutant *K-ras* is found in 80% or more of mouse lung adenomas and adenocarcinomas with mutations of codons 12 and 61 reported in both susceptible and resistant strains.[21 22] In hybrid animals, mutation occurs at a much higher frequency in susceptible alleles than in resistant alleles.[23] A bioassay in the same study showed that introduction of the mutant allele, derived from either the susceptible or resistant parent, was sufficient to transform NIH 3T3 cells. Recent studies have confirmed the strong association of the locus with cancer in mice and have shown that the homologous human locus, 12p12, is a major susceptibility locus in Italian and Japanese people.[17 24–26] The entire region between *K-ras* and *ITPR2* is homologous between mouse chromosome 6 and human chromosome 12, and human *K-ras* has been shown to share mutations associated with cancer at codons 12 and 61.[27] Some questions have remained. By what mechanism is the A/J allele of *K-ras* susceptible? Is there another gene in Pas1 that confers the propensity to mutation? Perhaps is it the *Par2* gene on chromosome 18, which has been shown to modulate the effect of Pas1? According to the present SNP analysis of Pas1, the *protein tyrosine phosphatase* and the *basic helix-loop-helix B3* genes are also candidates. Additional functional analysis of

the allelic differences in cell transformation and carcinogenesis should provide more definitive answers.

Pas2 QTL is located at the H-2 locus the haplotypes of which correlate with the incidence and multiplicity of induction of mouse lung tumour. The primary candidates for this region are *Notch4* and *Crebl1*. *Notch4* has been shown to be susceptible to viral activation, which can transform cells and drive them to a highly invasive phenotype.[11] This gene was also a candidate in previous expression studies. *Crebl1*, a SNPs binding protein, has been implicated in transcriptional activation and may contribute to the acquisition of metastatic phenotypes.[13]

Pas 3 was first described by Devereux *et al*[28] as flanked by D19MIT42 and D19MIT19. They used linkage analysis of N-ethyl-N-nitrosourea treated (A/J × C57BL/6J) Fl × C57BL/6J backcross progeny. This finding was later confirmed by Festing *et al*[29] with urethane treated (A/J × C57BL/6J) F2 mice. The main candidate in the Pas3 region is *Foxd4*. *Foxd4* is a forkhead transcription factor. Forkhead transcription factors are involved in regulating cell cycle progression and cell death. Long term Forkhead activation causes a sustained but reversible inhibition of proliferation without a marked increase in apoptosis.[14]

Finally, Pas4 was mapped by Festing *et al*.[29] One of the candidates is *TGFβRII*, the expression of which is reduced in urethane induced pulmonary adenoma in A/J mice. Loss of type 2 receptor results in a loss of induction of apoptosis.[30] Loss of sensitivity to *TGFβ* is a hallmark of invasiveness in some cancers.[31] However, no SNP between A/J and B6 was found, suggesting that the basis for its candidacy as Pas4 may be intronic or intergenic.

In this study, we also used RT-PCR to corroborate the results for six of seven candidate genes identified by allelic specific differential expression in our earlier study (fig 3).[7] Other work by us with others showed similar corroboration between RT-PCR and oligonucleotide arrays.[32 33] This strengthens the case for these genes potentially to play a part in predisposition of lung tumorigenesis in mice. However, the differences between these assays in reported fold change suggests that array results should continue to be taken cautiously and RT-PCR validation of the most important results should continue. Replication of arrays on multiple animals is also recommended. The study on gene expression complements the present study on SNPs in two ways. Firstly, differences in gene expression may not be the result of SNPs, so the gene expression based candidates bring in new information. Secondly, SNPs can have full effect without any change in gene expression whatsoever. So the sets of candidates are not mutually exclusive, nor are they subsets. Instead they are complementary.

In summary, we have described several candidates for the Pas and Par QTLs using the SNP analysis and microarrays. These candidates should accelerate identification of susceptibility genes to mouse lung tumour and potential functional interactions among different Pas and Par QTLs. This and our previous study show the synergies between classical genetic studies and genomic tools such as the SNP database and microarrays in dissecting the genetic basis of susceptibility to lung tumour in mice.

. . . .

Authors' affiliations

WJ Lemon, CH Swinton, M Wang, N Berbari, Y Wang, M You, Division of Human Cancer Genetics, Ohio State University Comprehensive Cancer Center, 420 West 12th Avenue, Columbus, Ohio 43210, USA

Correspondence to: Dr M You, Manuel Tzagournis Medical Research Facility, Room 514, 420 West 12th Avenue, Columbus, Ohio 43210, USA; you-1@medctr.osu.edu

REFERENCES

1 **Sellers TA**, Bailey-Wilson JE, Elston RC, Wilson AF, Elston GZ, Ooi WL, Rothschild H. Evidence for mendelian inheritance in the pathogenesis of lung cancer. *J Natl Cancer Inst* 1990;**82**:1272–9.

2 **Sellers TA**, Chen PL, Potter JD, Bailey-Wilson JE, Rothschild H, Elston RC. Segregation analysis of smoking-associated malignancies: evidence for Mendelian inheritance. *Am J Med Genet* 1994;**52**:308–14.

3 **Heston WE**, Dunn TB. Tumor development in susceptible strain A and resistant strain L lung transplants. *J Natl Cancer Inst* 1951;**11**:1057.

4 **Herzog CR**, Lubet RA, You M. Genetic epigenetic alterations in mouse lung tumors: implications for cancer chemoprevention. *J Cell Biochem* 1997;**28**:49–63.

5 **Chen B**, Johanson L, Wiest JS, Anderson MW, You M. The second intron of the K-ras gene contains regulatory elements associated with mouse lung tumor susceptibility. *Proc Natl Acad Sci USA* 1994;**91**:1589–93.

6 **Schork NJ**, Nath SP, Lindpaintner K, Jacob HJ. Extensions to quantitative trait locus mapping in experimental organisms. *Hypertension* 1996;**28**:1104–11.

7 **Lemon WJ**, Bernert H, Sun H, Wang Y, You M. Identification of candidate lung cancer susceptibility genes in mouse using oligonucleotide arrays. *J Med Genet* 2002;**39**:644–55.

8 **Ruivenkamp CA**, van Wezel T, Zanon C, Stassen AP, Vlcek C, Csikos T, Klous AM, Tripodis N, Perrakis A, Boerrigter L, Groot PC, Lindeman J, Mooi WJ, Meijjer GA, Scholten G, Dauwerse H, Paces V, van Zandwijk N, van Ommen GJ, Demant P. Ptprj is a candidate for the mouse colon-cancer susceptibility locus Scc1 and is frequently deleted in human cancers. *Nat Genet* 2002;**31**:295–300.

9 **Azmi S**, Taneja R. Embryonic expression of mSharp-1/mDEC2, which encodes a basic helix-loop-helix transcription factor. *Mech Dev* 2002;**114**:181.

10 **Fujimoto K**, Shen M, Noshiro M, Matsubara K, Shingu S, Honda K, Yoshida E, Suardita K, Matsuda Y, Kato Y. Molecular cloning and characterization of DEC2, a new member of basic helix-loop-helix proteins. *Biochem Biophys Res Commun* 2001;**280**:164–71.

11 **Callahan R**, Raafat A. Notch signaling in mammary gland tumorigenesis. *J Mammary Gland Biol Neoplasia* 2001;**6**:23–36.

12 **Xie S**, Price JE, Luca M, Jean D, Ronai Z, Bar-Eli M. Dominant-negative CREB inhibits tumor growth and metastasis of human melanoma cells. *Oncogene* 1997;**15**:2069–75.

13 **Jean D**, Bar-Eli M. Regulation of tumor growth and metastasis of human melanoma by the CREB transcription factor family. *Mol Cell Biochem* 2000;**212**:19–28.

14 **Kops GJ**, Medema RH, Glassford J, Essers MA, Dijkers PF, Coffer PJ, Lam EW, Burgering BM. Control of cell cycle exit and entry by protein kinase B-regulated forkhead transcription factors. *Mol Cell Biol* 2002;**22**:2025–36.

15 **Kang Y**, Mariano JM, Angdisen J, Moody TW, Diwan BA, Wakefield LM, Jakowlew SB. Enhanced tumorigenesis and reduced transforming growth factor-beta type II receptor in lung tumors from mice with reduced gene dosage of transforming growth factor-beta1. *Mol Carcinog* 2000;**29**:112–26.

16 **Kim TK**, Mo EK, Yoo CG, Lee CT, Han SK, Shim YS, Kim YW. Alteration of cell growth and morphology by overexpression of transforming growth factor beta type II receptor in human lung adenocarcinoma cells. *Lung Cancer* 2001;**31**:181–91.

17 **Manenti G**, Stafford A, De Gregorio L, Gariboldi M, Falvella FS, Avner P, Dragani TA. Linkage disequilibrium and physical mapping of Pas1 in mice. *Genome Res* 1999;**9**:639–46.

18 **Festing MF**, Simpson EM, Davisson MT, Mobraaten LE. Revised nomenclature for strain 129 mice. *Mamm Genome* 1999;**10**:836.

19 **Simpson EM**, Linder CC, Sargent EE, Davisson MT, Mobraaten LE, Sharp JJ. Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice. *Nat Genet* 1997;**16**:19–27.

20 **Threadgill DW**, Yee D, Matin A, Nadeau JH, Magnuson T. Genealogy of the 129 inbred strains: 129/SvJ is a contaminated inbred strain. *Mamm Genome* 1997;**8**:390–3.

21 **Gariboldi M**, Manenti G, Canzian F, Falvella FS, Radice MT, Pierotti MA, Della Porta G, Binelli G, Dragani TA. A major susceptibility locus to murine lung carcinogenesis maps on chromosome 6. *Nat Genet* 1993;**3**:132–6.

22 **You M**, Candrian U, Maronpot RR, Stoner GD, Anderson MW. Activation of the Ki-ras protooncogene in spontaneously occurring and chemically induced lung tumors of the strain A mouse. *Proc Natl Acad Sci USA* 1989;**86**:3070–4.

23 **You M**, Wang Y, Stoner G, You L, Maronpot R, Reynolds SH, Anderson M. Parental bias of Ki-ras oncogenes detected in lung tumors from mouse hybrids. *Proc Natl Acad Sci USA* 1992;**89**:5804–8.

24 **Manenti G**, De Gregorio L, Pilotti S, Falvella FS, Incarbone M, Ravagnani F, Pierotti MA, Dragani TA. Association of chromosome 12p genetic polymorphisms with lung adenocarcinoma risk and prognosis. *Carcinogenesis* 1997;**18**:1917–20.

25 **Manenti G**, Nomoto T, De Gregorio L, Gariboldi M, Stefania Falvella F, Nagao M, Dragani TA. Predisposition to lung tumorigenesis. *Toxicol Lett* 2000;**112–113**:257–63.

26 **Dragani TA**, Hirohashi S, Juji T, Kawajiri K, Kihara M, Ono-Kihara M, Manenti G, Nomoto T, Sugimura H, Genka K, Yokota J, Takahashi T, Mitsudomi T, Nagao M. Population-based mapping of pulmonary adenoma susceptibility 1 locus. *Cancer Res* 2000;**60**:5017–20.

27 **Abrams ES**, Murdaugh SE, Lerman LS. Comprehensive screening of the human KRAS2 gene for sequence variants. *Genes Chrom Cancer* 1993;**6**:73–85.

28 **Devereux TR**, Wiseman RW, Kaplan N, Garren S, Foley JF, White CM, Anna C, Watson MA, Patel A, Jarchow S, Maronpot RR, Anderson MW. Assignment of a locus for mouse lung tumor susceptibility to proximal chromosome 19. *Mamm Genome* 1994;**5**:749–55.

29 **Festing MFW**, Lin L, Devereux TR, Gao F, Yang A, Anna CH, White CM, Malkinson AM, You M. At least four loci associated with susceptibility to the chemical induction of lung adenomas in mice. *Genomics* 1998;**53**:129–36.

30 **de Jonge RR**, Garrigue-Antar L, Vellucci VF, Reiss M. Frequent inactivation of the transforming growth factor beta type II receptor in small-cell lung carcinoma cells. *Oncol Res* 1997;**9**:89–98.

31 **Teicher BA**. Malignant cells, directors of the malignant process: role of transforming growth factor-β. *Cancer Metastasis Rev* 2001;**20**:133–43.

32 **Huang Y**, Prasad M, Lemon WJ, Hampel H, Wright FA, Kornacker K, LiVolsi V, Frankel W, Kloos RT, Eng C, Pellegata NS, de la Chapelle A. Gene expression in papillary thyroid carcinoma reveals highly consistent profiles. *Proc Natl Acad Sci USA* 2001;**98**:15044–9.

33 **Virtaneva K**, Wright FA, Tanner SM, Yuan B, Lemon WJ, Caligiuri MA, Bloomfield CD, de La Chapelle A, Krahe R. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. *Proc Natl Acad Sci USA* 2001;**98**:1124–9.