Arsenic-related DNA copy-number alterations in lung squamous cell carcinomas

VD Martinez[*,1], **TPH Buys**[2], **M Adonis**[1], **H Benítez**[3], **I Gallegos**[4], **S Lam**[2], **WL Lam**[2] and **L Gil**[1]

[1]Department of Molecular and Cell Biology, Faculty of Medicine, Biomedical Sciences Institute, University of Chile, Independencia 1027, Santiago 8380453, Chile; [2]Department of Integrative Oncology, BC Cancer Research Centre, 675 West 10th Avenue, Vancouver, British Columbia, Canada V5Z 1L3; [3]Unidad de Anatomía Patológica, Hospital Regional de Antofagasta, Av. Argentina 1962, Antofagasta 1270001, Chile; [4]Departamento de Anatomía Patológica, Hospital Clínico Universidad de Chile, Av. Independencia 1027, Santiago 8380453, Chile

BACKGROUND: Lung squamous cell carcinomas (SqCCs) occur at higher rates following arsenic exposure. Somatic DNA copy-number alterations (CNAs) are understood to be critical drivers in several tumour types. We have assembled a rare panel of lung tumours from a population with chronic arsenic exposure, including SqCC tumours from patients with no smoking history.
METHODS: Fifty-two lung SqCCs were analysed by whole-genome tiling-set array comparative genomic hybridisation. Twenty-two were derived from arsenic-exposed patients from Northern Chile (10 never smokers and 12 smokers). Thirty additional cases were obtained for comparison from North American smokers without arsenic exposure. Twenty-two blood samples from healthy individuals from Northern Chile were examined to identify germline DNA copy-number variations (CNVs) that could be excluded from analysis.
RESULTS: We identified multiple CNAs associated with arsenic exposure. These alterations were not attributable to either smoking status or CNVs. DNA losses at chromosomes 1q21.1, 7p22.3, 9q12, and 19q13.31 represented the most recurrent events. An arsenic-associated gain at 19q13.33 contains genes previously identified as oncogene candidates.
CONCLUSIONS: Our results provide a comprehensive approach to molecular characteristics of the arsenic-exposed lung cancer genome and the non-smoking lung SqCC genome. The distinct and recurrent arsenic-related alterations suggest that this group of tumours may be considered as a separate disease subclass.
British Journal of Cancer (2010) **103**, 1277–1283. doi:10.1038/sj.bjc.6605879   www.bjcancer.com
Published online 14 September 2010
© 2010 Cancer Research UK

Keywords: arsenic; lung squamous cell carcinoma; never smokers; copy-number alterations; copy-number variation; CGH

Arsenic is a well-established skin, bladder, and lung carcinogen, and ~40 million individuals worldwide are regularly exposed to it, mainly through drinking water (Nordstrom, 2002; IARC (International Agency for Research on Cancer), 2004). In the Northern Chilean region of Antofagasta, the population (~300 000 inhabitants) was chronically exposed to arsenic concentrations as high as $870 \mu g l^{-1}$ from 1958 through 1970, with some towns still registering concentrations of $600 \mu g l^{-1}$ by the year 2000 (Smith et al, 1998). This has occurred despite international guidelines placing the maximum tolerable arsenic exposure at $10 \mu g l^{-1}$ (IARC, 2004).

Lung cancer (LC) is the leading cause of cancer-related mortality worldwide (Jemal et al, 2006). Lung squamous cell carcinomas (SqCCs) in particular are strongly tied to exposure to tobacco smoke (Herbst et al, 2008). Lung cancer is also the main cause of deaths due to arsenic exposure, with this carcinogen acting as the major aetiological agent in cancers that occur in never smokers (Mead, 2005). Relative to other histological subtypes of LC, lung SqCCs are also strongly associated with arsenic exposure (Guo et al, 2004). Concordantly, rates for this disease in Antofagasta are the highest in Chile; SqCCs account for approximately two-thirds of all LC cases in this region (Servicio de Salud Antofagasta, 2000).

By contrast, SqCCs account for ~20% of LCs in North America, where arsenic exposure is absent or limited (Horner et al, 2009). Since the 1990s, adenocarcinomas (AdCs) have become the predominant LC cell type especially among never smokers (Samet et al, 2009). It has become increasingly difficult to obtain SqCCs from never smokers to study the mechanism of LC development unrelated to tobacco smoking to understand alternative pathways to carcinogenesis. Therefore, our arsenic-exposed never smokers represent a unique resource to study the molecular pathology of lung SqCCs.

The mechanism by which arsenic causes cancer is still under investigation. It has been proposed that both genetic and epigenetic processes are involved (Salnikow and Zhitkovich, 2008). Interestingly, specific DNA copy-number alterations (CNAs) have been described in bladder cancer cases from different populations exposed to arsenic in drinking water, including some from Northern Chile (Moore et al, 2002; Hsu et al, 2008). These results suggest that these kinds of alterations are likely to have a central role in the development and progression of some arsenic-related cancers.

The CNAs are key events in tumour progression (Hanahan and Weinberg, 2000). These changes have been widely identified among phenotypically normal individuals (i.e., germline DNA copy-number variations (CNVs)) (Redon et al, 2006; Wong et al, 2007). The CNAs, as well as CNVs, include gains/losses of DNA segments, which can potentially disrupt regulation of gene

expression by dosage or positional effects (Feuk et al, 2006). These alterations have been extensively described in different LC subtypes (Pei et al, 2001; Powell et al, 2003; Garnis et al, 2005). Additionally, non-smoking aetiological agents, such asbestos, exhibit specific CNAs (Nymark et al, 2006; Kettunen et al, 2009). To date, there have been no attempts to identify CNAs in lung tumours in which arsenic is thought to be the main aetiological factor. Moreover, the non-smoking lung SqCC genome has not been fully evaluated. This is likely a result of the scarcity of such cases or the difficulties inherent to discriminating arsenic-related alterations from those induced by other common lung carcinogens or polymorphic variations.

In this report, we defined arsenic-associated LC CNAs by applying a whole-genome tiling-path array-based comparative genomic hybridisation (CGH) platform to a rare panel of arsenic-exposed lung SqCCs collected from a Northern Chilean population (Watson et al, 2007). We compare the results for these specimens with those derived from a panel of lung SqCCs from an unrelated population without known exposure to arsenic. Smoking-associated alterations were subtracted by in-group comparisons. Germline CNVs associated with the Northern Chilean population were subtracted from analysis by evaluation of DNA isolated from the blood of unrelated, healthy patients from the same region. This work represents a comprehensive genomic analysis of DNA copy-number changes in LCs in which arsenic is suspected to be the primary aetiological factor. It also provides molecular insights into the non-smoking lung SqCC genome.

MATERIALS AND METHODS

Accrual of clinical LC samples and DNA extraction from tumours

A total of 52 cases of lung SqCC were analysed in this study. Twenty-two formalin-fixed paraffin embedded lung SqCC specimens were collected from the Regional Hospital of Antofagasta, Northern Chile (10 corresponding to never smokers and 12 to smokers). Another panel of 30 lung SqCC tumours from patients with a history of tobacco use and no known exposure to arsenic was processed in a similar manner (this latter group was diagnosed at the British Columbia Cancer Agency in Vancouver, Canada). Demographics, including age, sex, area of residence and so on, were obtained for each case (Supplementary Material 1), with histological diagnosis confirmed by expert pathology review. 'Never Smoker' classifications were made based on information from medical records using previously defined criteria (Sun et al, 2007). Pathology review served to determine tumour cell content for each case. Tumour cells were isolated from tissue cross-sections by either manual or laser-assisted microdissection depending on tissue heterogeneity. Microdissected tissue was placed in an SDS solution with proteinase K at 48°C and spiked with additional enzyme twice a day for 72 h. DNA was extracted by a standard phenol: chloroform method, and DNA concentrations were determined using a NanoDrop Spectrophotometer (NanoDrop Technologies, Houston, TX, USA) (Sambrook et al, 1989). An independent group of 22 samples of peripheral blood derived from phenotypically normal individuals who were also exposed to arsenic (also from Northern Chile) were processed in the same form. Demographics for these samples were obtained through personal interview (after reading and signing an informed consent) and are also shown in Supplementary Material 1. The Ethics Committee of the Faculty of Medicine of the University of Chile (case ID: 085-2008) approved the study and the analysis of both formalin-fixed paraffin embedded samples and interview data.

Array CGH experiments

Genome alteration profiles were generated using the sub-megabase resolution tiling-set rearray (SMRTr) platform, which is comprised

of >26 000 duplicate-spotted BAC clones that span the whole human genome (Watson et al, 2007). Briefly, 200 ng of tumour/blood or a single-male reference DNA samples was labelled with 2 μl (2 nmol) of cyanine-3 dCTP or cyanine-5 dCTP (Perkin Elmer Life Sciences Inc., Boston, MA, USA), respectively. Labelled probes were incubated for 18 h at 37°C, purified, precipitated with Cot-1 DNA (Invitrogen, Burlington, Ontario, Canada), and resuspended in hybridisation buffer. A 10-min denaturing step at 85°C followed and then repetitive sequences were blocked at 45°C for 1 h. Probe mixtures were then applied to the SMRTr array surface and allowed to competitively hybridise for 40 h at 45°C. Following hybridisation, slides underwent agitating 45°C washes with 0.1 × saline sodium citrate, 0.1% SDS, and then rinses with 0.1 × saline sodium citrate followed by drying with centrifugation.

Imaging and data processing

Post-hybridisation arrays were analysed as previously described (Watson et al, 2007). Slides were scanned using a charge-coupled device-based imaging system (arrayWoRx eAuto, Applied Precision). The spot signal intensity ratios from images were calculated using the SoftWoRx Tracker Spot Analysis (Applied Precision, Issaquah, WA, USA). Data were normalised using a previously described stepwise framework (Khojasteh et al, 2005). Log2 ratios were plotted using custom SeeGH software (Chi et al, 2008). BAC clones with variance >0.01 and signal-to-noise ratio <10 were excluded from further analysis. Additionally, samples with a global s.d. >0.15 across autosomal BAC clones were not considered.

DNA CNA detection

The CNAs were defined using a previously described heuristic algorithm (Jong et al, 2004). Frequency of alteration at each clone on the array was then calculated. Clones uninformative in >10% of samples were excluded from further experiments. Analysis was also restricted to autosomes, with any differences based on sex subtracted from further comparisons. DNA segments showing log2 ratios >0.2 for DNA gains and <−0.2 for losses were considered as significant. Significant differences in CNA frequency were evaluated using the Fisher's exact test. P-values were adjusted using the Benjamini–Hochberg method in order to control the false discovery rate (Benjamini and Hochberg, 1995). An adjusted P-value <0.05 was used as a significance threshold. Single clone alterations were excluded from further analysis, ensuring that multiple, neighbouring clones spanned candidate regions. Finally, visual inspection of normalised data plotted in SeeGH software validated those identified segments as CNAs.

Gene content

Through a combined analysis of all genome data for these tumours, we defined minimal common regions (MCRs) of recurring DNA gain and loss. This approach has been widely used for CGH data and is useful for refining our list of genes potentially affected (Tonon et al, 2005; Lockwood et al, 2008). The MCRs were specifically defined as the minimal region of DNA spanned by two or more BAC clones exhibiting the higher frequency of DNA gain/loss across sample set. Gene content and genomic locations at MCRs were obtained using the Galaxy Platform (http://main.g2.bx.psu.edu/), based on March 2006 human genome reference sequence (NCBI Build 36.1) (Giardine et al, 2005).

RESULTS

Differences in CNA patterns among arsenic-exposed and non-exposed lung SqCCs

We first independently characterised commonly occurring CNAs on arsenic-exposed and non-exposed groups using a tiling-path

whole-genome array CGH platform. The CNAs associated with arsenic exposure were defined using the framework shown in Figure 1. First, we sought to identify differences between exposed and non-exposed groups. To that end, we compared genomic profiles from 22 tumours taken from individuals with known arsenic exposure in Northern Chile (12 smokers, 10 never smokers) against profiles from 30 lung SqCCs obtained from North America (none of this latter panel had any documented exposure to arsenic). Results are shown in Figure 2. On the basis of the Fisher's exact test – and using a Benjamini–Hochberg-adjusted P-value cutoff of <0.05 – we identified 68 clones as being different between these groups. These differences could be related to arsenic exposure, smoking status, or CNVs.

Identification of arsenic-related CNAs

To filter out CNAs associated with smoking, genomic profiles from within the Chilean subset of SqCC tumours were next compared.

Although all of these cases were known to have arsenic exposure, 10 were never smokers and 12 were current/former smokers. First, we performed a whole-genome comparison between these groups. Given what is known about recurring genomic alterations in recurring lung SqCCs, the low frequency of DNA gains at chromosome 3q among arsenic-exposed never smokers was of particular interest (Figure 3).

We next focused our analysis on the 68 clones defined as differentially altered above. The rationale for this approach was that shared alterations between these two groups should not be linked to tobacco smoke (i.e., shared alterations would more likely be associated with either arsenic exposure or CNVs tied to the population of Antofagasta, Chile). With this approach, 58 BAC clones did not show significant differences (considering a significance threshold for Fisher's exact test at $P<0.05$). These were retained for further analysis.

With 58 clones now tied to either arsenic exposure or regional germline CNVs, we next attempted to filter data to only those
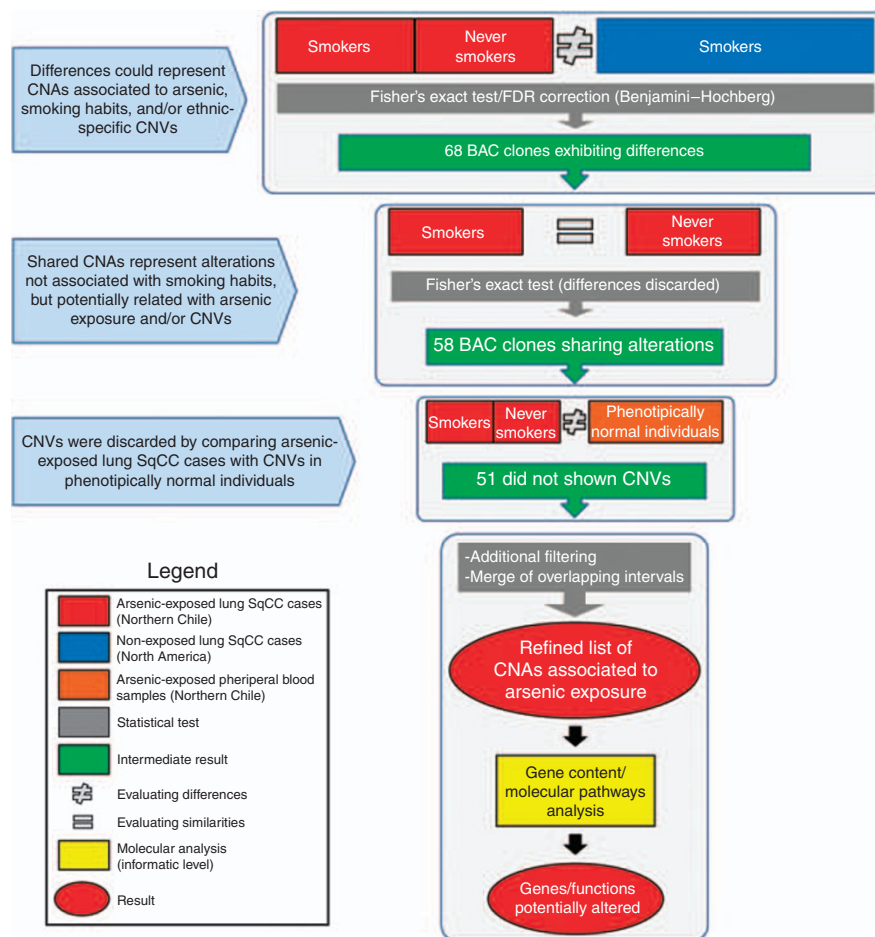
Figure 1 Study framework for identifying DNA alterations associated to arsenic exposure. To discover copy-number alterations (CNAs) associated with arsenic exposure, we initially screened for differences in alteration frequencies between lung squamous cell carcinoma (SqCC) cases from arsenic-exposed regions (both smokers and never smokers from Northern Chile, $n = 22$) and non-exposed regions (from North America, $n = 30$). Comparisons were performed using the Fisher's exact test, with P-value corrected for false-discovery rate (FDR). As a result, 68 bacterial artificial chromosome (BAC) clones were considered to be differentially altered between these groups (corrected P-value <0.05). Differences could be significantly associated with arsenic exposure, smoking status, or copy-number variations (CNVs) at this stage. In order to filter out smoking-related alterations, cases from arsenic-exposed smokers and never smokers were next compared. This analysis was restricted to those regions defined by the first comparison (i.e., 68 BAC clones). Shared alterations were kept for downstream analysis as these would be unlikely to be related to smoking. Fifty-eight clones shared CNAs status between these arsenic-exposed lung SqCC groups. Next, we filtered out germline CNVs associated with the population in Northern Chile: the status of the 58 BAC clones defined above was evaluated in genome profiles generated from peripheral blood samples obtained from an independent group of 22 phenotipically normal individuals who were exposed to arsenic (also from Northern Chile). Any clones identified as CNVs in this normal population were removed from further analysis. To make our analysis even more robust, we restricted our regions of interest to those alterations that (1) involved multiple, overlapping array clones and (2) showed log2 ratios >0.2 or <-0.2 for gains and losses, respectively. Eight unique regions of alteration were found to be significantly associated with arsenic exposure. Genes mapping to these regions (based on NCBI Build 36.1) were then used for molecular pathway analysis and detailed literature reviews to determine their relevance to cancer and arsenic metabolism.

(−) 50%   (+) 50%
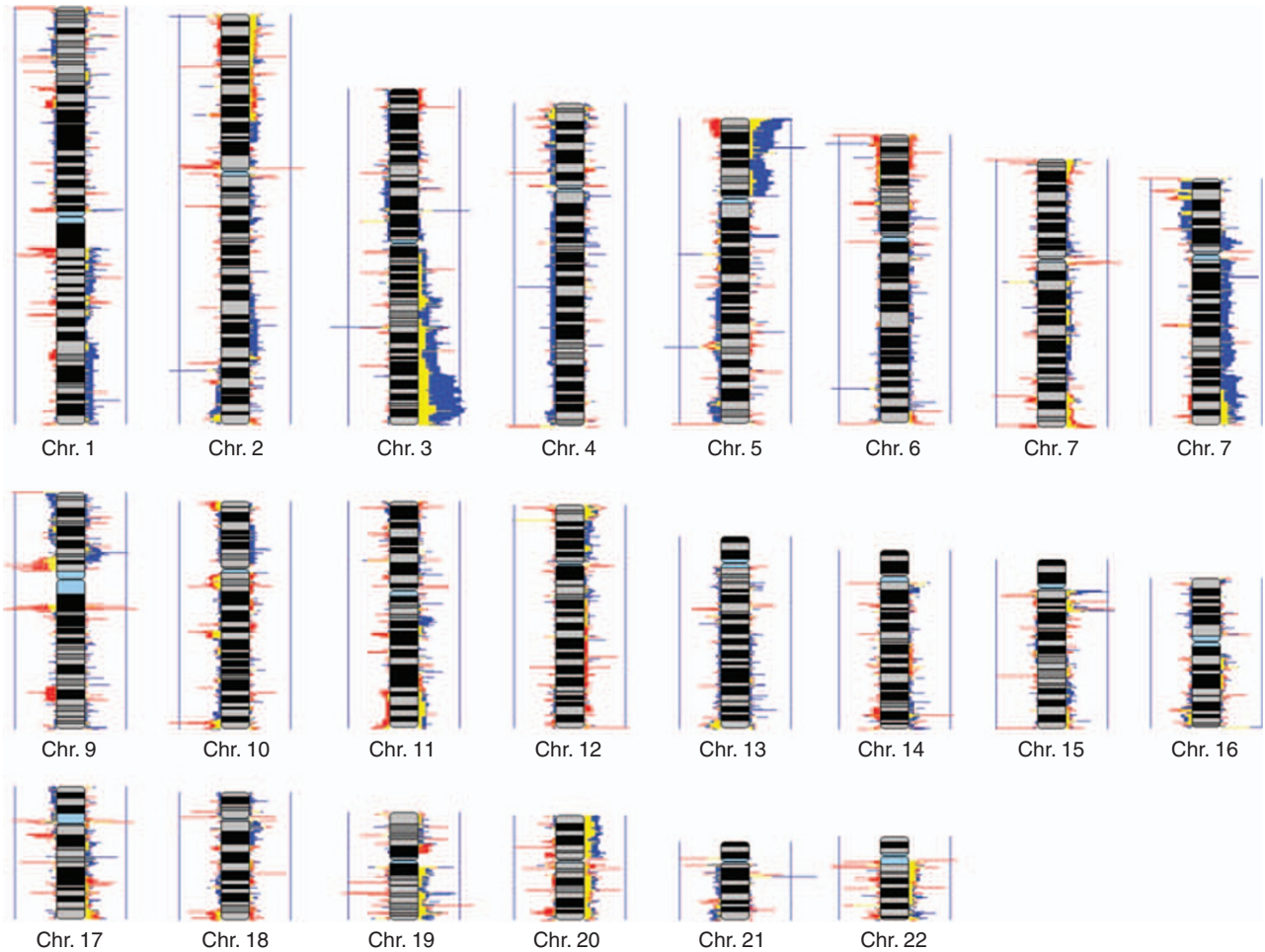
Figure 2 Comparison of DNA alteration frequency for lung squamous cell carcinoma (SqCC) from arsenic-exposed *vs* non-exposed patients. Genome profiles were generated for 52 lung SqCC biopsies derived from 22 arsenic-exposed smoker and never-smoker patients from Northern Chile (red), and 30 current and ex-smoker North American patients without known arsenic exposure or non-exposure (blue). Data from each patient were aligned and frequency of DNA gain/loss for each bacterial artificial chromosome (BAC) clone was calculated across the complete sample set. Frequency of alteration results for exposed and non-exposed SqCC cases have been overlaid in this figure. Results in yellow denote a region of overlapping alteration status in both groups. The magnitude of red and blue bars represents percentage of samples exhibiting corresponding alteration (0–100%, with blue vertical lines representing 50% frequency). DNA gains and losses are represented to the right and left of each chromosome, respectively. Analysis was restricted to autosomes, with any differences based on sex subtracted from further analysis (see Materials and Methods). Chr., chromosome.

alterations associated with arsenic exposure. First, blood was drawn from an independent, unrelated group of 22 individuals from the same region of Northern Chile. None of these individuals had any recorded incidence of LC or other neoplastic processes and all had been exposed to arsenic in drinking water. DNA isolated from these individuals was then profiled and analysed in the same manner described above. Of the 58 clones defined after our first two analyses, only 7 were detected as CNVs in this independent sample set (mainly at chromosome 1). This left 51 BAC clones as CNAs associated with arsenic exposure. To further restrict our analysis to the most robust differences, only regions spanned by two or more overlapping BAC clones that showed log2 ratios >0.2 or <-0.2 were considered as regions of alteration (gains and losses, respectively). This reduced the list to 32 BAC clones and 8 unique regions of CNAs that occurred at higher frequency in arsenic-exposed lung SqCC cases from Northern Chile. These CNAs corresponded to seven regions of DNA loss (chromosome bands 1q21.1, 2p11.2, 2p11.1, 7p22.3, 9q12, 15q26.3, and 19q13.31) and one region of DNA gain at 19q13.33 (Table 1).

Finally, we investigated the specific genes associated with the arsenic-related CNAs described above. A total of 31 genes were

present at MCRs of alteration (Table 1). More than half of the identified genes were located at two regions of arsenic-specific CNAs detected at 19q. DNA loss at 19q13.31 (which was not detected among non-exposed or as CNVs) contains eight genes, which belong to the family of human pregnancy-specific glycoproteins. On the other hand, some cancer-related genes, such as *POLD1* (DNA polymerase δ 1, catalytic subunit), *SPIB* (Spi-B transcription factor), and *NR1H2* (nuclear receptor subfamily 1, group H, member 2) were located at 19q13.33, the single region of DNA gain associated with arsenic-exposed cases.

DISCUSSION

Tobacco exposure is unquestionably the main aetiological factor in LC. That said, arsenic ingestion through drinking water also represents a critical cause of LC – particularly lung SqCCs – (Ferreccio *et al*, 2000). In this study, we performed a comprehensive high-resolution characterisation of CNAs in lung SqCC with different records of arsenic exposure. In addition, we have screened the non-smoking lung SqCC genome. By this approach,
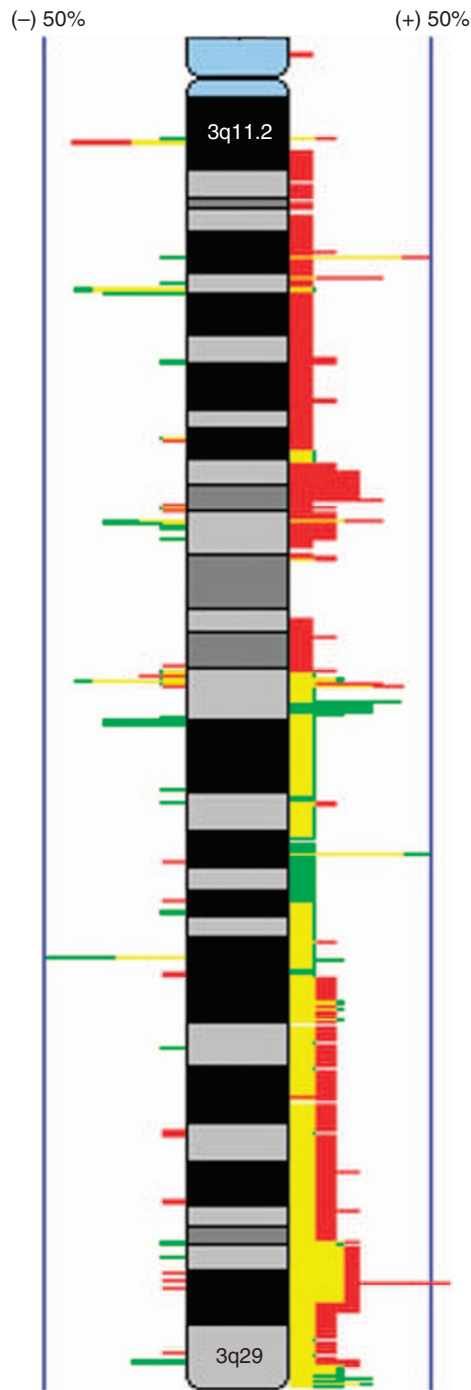
Figure 3 Comparison of copy-number alterations (CNAs) frequency at chromosome region 3q for lung squamous cell carcinomas (SqCCs) derived from arsenic-exposed smokers and never smokers. Genome profiles for 22 cases of arsenic-exposed lung SqCC biopsies (12 from current and ex-smokers and 10 from never smokers) were compared at chromosome region 3q and visualised using SeeGH software. Frequency of CNAs results for smokers and never smokers have been overlaid in this figure. Results for smokers are represented in red; while results for never smokers are shown in green (yellow denotes a region exhibiting similar alteration status in both groups). The magnitude of green and red bars represents percentage of samples exhibiting corresponding alteration (0–100%, with blue vertical lines representing 50% frequency). DNA gains and losses are displayed to the right and left of each chromosome, respectively.

we have identified recurrent segmental DNA gains and losses in lung SqCC following arsenic intake through drinking water.

First, we noted a very low frequency of DNA gains at the terminal portion of chromosome arm 3q for lung SqCCs from arsenic-exposed patients. This frequency was subsequently found to be particularly lower among arsenic-exposed never smokers (arsenic-exposed smokers showed a similar frequency of 3q gain as non-exposed cases). Indeed, <30% of cases from arsenic-exposed never smokers exhibited DNA gains at 3q and only at selected locations (data not shown), which is similar to the frequency described for lung AdCs (Luk et al, 2001; Pei et al, 2001). This result is remarkable, as this alteration – especially between 3q26-q29 – is the most widely reported genomic alteration associated with lung SqCC tumours and cell lines (Tonon et al, 2005; Garnis et al, 2006). Our results suggest that additional alterations outside of increased 3q dosage may drive the emergence of lung SqCCs, particularly where the aetiological basis for disease is not tobacco smoke.

Another key finding of this work was the particular association of DNA losses with arsenic-exposed lung SqCC. This result is concordant with previous findings showing that arsenic can induce multiple large deletions through the creation of reactive oxygen species (Concha et al, 1998). DNA losses at 1q21.1, 7p22.3, and 19q13.31 were only detected among arsenic-exposed cases, suggesting that these alterations could have a role in arsenic-induced lung neoplasia. We also detected a ∼300 kb deletion at 9q12 in nearly two-thirds of samples from arsenic-exposed patients. This result is striking given that chromosome 9q deletion has previously been reported at high incidence in bladder tumours from never-smoker patients in Argentina and Chile who were exposed to high levels of arsenic (Moore et al, 2002). One of the genes at 9q12, FOXD4L6, belongs to the forkhead box (Fox) gene family. Although neither FOXD4L6 nor other members of the FoxD gene subfamily have been tied to neoplastic processes, several other Fox genes have been linked to tumourigenesis and cancer progression (including LC) (Kim et al, 2006). This suggests that DNA loss at 9q12, particularly of the FOXD4L6 gene, could be relevant where arsenic exposure is a critical cancer-causing mechanism.

The only DNA gain associated with arsenic-exposed cases occurred at 19q13.33, and this alteration was detected for 68.2% of cases. Interestingly, this alteration occurred at an even higher rate in tumours from arsenic-exposed never smokers (90% of cases). This region has previously been described as recurrently altered in lung AdCs, including as a lung AdCs susceptibility locus (Yanagitani et al, 2003; Tonon et al, 2005). Moreover, it has been proposed that allelic imbalance in this region is essential in lung AdCs arising in non-smoking patients (Wong et al, 2002). Genes mapping to 19q13.33 have previously been described as altered by changes in DNA dosage. For example, the SPIB gene has been described as gained in lung AdCs (Tonon et al, 2005). SPIB has also been shown to be gained and overexpressed in B-cell-like diffuse large B-cell lymphoma (Lenz et al, 2008). This region also contains the POLD1 gene, which codes for the proofreading domain of the DNA polymerase δ complex. POLD1 participates in DNA single-strand break repair process and has been shown to increase genomic instability and incidence of epithelial tumours in mice (Goldsby et al, 2002; Parsons et al, 2007; Venkatesan et al, 2007). On the basis of this activity, disruption of POLD1 function could have a significant role in lung tumourigenesis. NR1H2, which also maps to this region, has been related with spleen lymphoid hyperplasia in mouse and could also have a role in the emergence of this particular subset of lung SqCCs (Bensinger et al, 2008). These data suggest that gain of 19q13.33 may help drive lung SqCCs in arsenic-exposed individuals via multiple mechanisms, including activation of oncogenes, increasing genomic instability, and disrupted DNA repair. Analysis of a larger sample set and cell model systems is clearly needed to more precisely delineate the

Genetics and Genomics

Table 1 Specific DNA CNAs in arsenic-exposed lung SqCC cases

Cytoband	Start (Mb)	End (Mb)	Size (Mb)	No. of overlapping BAC clones in MCR	Type of CNA	% of samples exhibiting corresponding type of CNA			Genes
						As (%)[a]	No As (%)[b]	CNV (%)[c]	
1q21.1	146.8	147.2	0.4	2	L-As	50.0	ND	ND	*NBPF15, NBPF16, PPIAL4E, LOC645166*
2p11.2	89.6	89.9	0.3	3	L-As	50.0	ND	4.5	No known genes in this region
2p11.1	91.0	91.5	0.4	5	L-As	50.0	ND	4.5	*LOC654342, GGT8P*
7p22.3	0.0	0.3	0.3	6	L-As	50.0	ND	ND	*FAM20C*
9q12	68.4	68.8	0.3	2	L-As	63.6	10.0	4.5	*LOC440896, FOXD4L6, CBWD6, ANKRD20A4*
15q26.3	100.1	100.3	0.2	2	L-As	50.0	3.3	4.5	*OR4F6, OR4F15, OR4F4*
19q13.31	47.9	48.3	0.4	3	L-As	50.0	ND	ND	*PSG3, PSG8, PSG18, PSG1, PSG6, PSG7, PSG11, PSG2*
19q13.33	55.6	55.7	0.2	3	G-As	68.2	10.0	ND	*NAPSA, NR1H2, POLD1, SPIB, MYBPC2, FAM71E1, C19orf63, JOSD2, ASPDH*

Abbreviations: As = arsenic; CNA = copy number alteration; CNV = copy-number variation; G = gain; L = loss; MCR = minimal common region; ND = not determined; SqCC = squamous cell carcinoma. [a]Arsenic exposed. [b]Non-exposed to arsenic. [c]DNA CNVs evaluated in arsenic-exposed individuals.

molecular basis for both arsenic-induced LC and SqCC in never smokers. Unfortunately, such studies are largely precluded by the relative rarity of appropriate specimens. In this context, the results we present here are a key first step in analysing this subset of lung tumours.

In sum, we have used high-resolution genomic profiling to identify arsenic-related CNAs in lung SqCC. This has provided insights into the molecular mechanisms of both arsenic-induced cancers and non-smoking lung SqCC. Many genomic loci that we identified as significantly associated with this lung tumour subset have been previously implicated in cancer processes, including candidate genes associated specifically in other arsenic-induced tumours and other LC subtypes. As we analysed DNA samples from formalin-fixed paraffin embedded biopsies, further molecular analysis to establish functional significance is not possible. However, biological information obtained from this exceptionally uncommon subset of arsenic-exposed lung SqCC cases (especially those arisen in never smokers) is valuable. Future efforts will be focused on functional consequences of arsenic-related CNAs, specifically on genes located at chromosome 19. Taken together,

our data suggest that lung SqCC in arsenic-exposed individuals could represent a molecularly distinct form of lung SqCCs. Given a unique set of underlying genomic changes, distinct approaches to treatment may be appropriate for this patient population.

REFERENCES

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* **57**: 289–300

Bensinger SJ, Bradley MN, Joseph SB, Zelcer N, Janssen EM, Hausner MA, Shih R, Parks JS, Edwards PA, Jamieson BD, Tontonoz P (2008) LXR signaling couples sterol metabolism to proliferation in the acquired immune response. *Cell* **134**: 97–111

Chi B, deLeeuw RJ, Coe BP, Ng RT, MacAulay C, Lam WL (2008) MD-SeeGH: a platform for integrative analysis of multi-dimensional genomic data. *BMC Bioinformatics* **9**: 243

Concha G, Vogler G, Lezcano D, Nermell B, Vahter M (1998) Exposure to inorganic arsenic metabolites during early human development. *Toxicol Sci* **44**: 185–190

Ferreccio C, Gonzalez C, Milosavjlevic V, Marshall G, Sancha AM, Smith AH (2000) Lung cancer and arsenic concentrations in drinking water in Chile. *Epidemiology* **11**: 673–679

Feuk L, Marshall CR, Wintle RF, Scherer SW (2006) Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum Mol Genet* **15**(Spec. No. 1): R57–R66

Garnis C, Davies JJ, Buys TP, Tsao MS, MacAulay C, Lam S, Lam WL (2005) Chromosome 5p aberrations are early events in lung cancer: implication of glial cell line-derived neurotrophic factor in disease progression. *Oncogene* **24**: 4806–4812

Garnis C, Lockwood WW, Vucic E, Ge Y, Girard L, Minna JD, Gazdar AF, Lam S, MacAulay C, Lam WL (2006) High resolution analysis of non-small cell lung cancer cell lines by whole genome tiling path array CGH. *Int J Cancer* **118**: 1556–1564

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* **15**: 1451–1455

Goldsby RE, Hays LE, Chen X, Olmsted EA, Slayton WB, Spangrude GJ, Preston BD (2002) High incidence of epithelial cancers in mice deficient for DNA polymerase delta proofreading. *Proc Natl Acad Sci USA* **99**: 15560–15565

Guo HR, Wang NS, Hu H, Monson RR (2004) Cell type specificity of lung cancer associated with arsenic ingestion. *Cancer Epidemiol Biomarkers Prev* **13**: 638–643

Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* **100**: 57–70

Herbst RS, Heymach JV, Lippman SM (2008) Lung cancer. *N Engl J Med* **359**: 1367–1380

Horner MJ, Ries LAG, Krapcho M, Neyman N, Aminou R, Howlader N, Altekruse SF, Feuer EJ, Huang L, Mariotto A, Miller BA, Lewis DR, Eisner MP, Stinchcomb DG (eds) (2009) *SEER Cancer Statistics Review, 1975–2006*. National Cancer Institute: Bethesda, MD

Hsu LI, Chiu AW, Pu YS, Wang YH, Huan SK, Hsiao CH, Hsieh FI, Chen CJ (2008) Comparative genomic hybridization study of arsenic-exposed and non-arsenic-exposed urinary transitional cell carcinoma. *Toxicol Appl Pharmacol* **227**: 229–238

IARC (International Agency for Research on Cancer) (2004) Tobacco smoke and involuntary smoking. *IARC Monogr Eval Carcinog Risks Hum* **83**: 1

Jemal A, Siegel R, Ward E, Murray T, Xu J, Smigal C, Thun MJ (2006) Cancer statistics, 2006. *CA Cancer J Clin* **56**: 106–130

Genetics and Genomics

Jong K, Marchiori E, Meijer G, Vaart AV, Ylstra B (2004) Breakpoint identification and smoothing of array comparative genomic hybridization data. *Bioinformatics* **20**: 3636–3637

Kettunen E, Aavikko M, Nymark P, Ruosaari S, Wikman H, Vanhala E, Salmenkivi K, Pirinen R, Karjalainen A, Kuosma E, Anttila S (2009) DNA copy number loss and allelic imbalance at 2p16 in lung cancer associated with asbestos exposure. *Br J Cancer* **100**: 1336–1342

Khojasteh M, Lam WL, Ward RK, MacAulay C (2005) A stepwise framework for the normalization of array CGH data. *BMC Bioinformatics* **6**: 274

Kim IM, Ackerson T, Ramakrishna S, Tretiakova M, Wang IC, Kalin TV, Major ML, Gusarova GA, Yoder HM, Costa RH, Kalinichenko VV (2006) The Forkhead Box m1 transcription factor stimulates the proliferation of tumor cells during development of lung cancer. *Cancer Res* **66**: 2153–2161

Lenz G, Wright GW, Emre NC, Kohlhammer H, Dave SS, Davis RE, Carty S, Lam LT, Shaffer AL, Xiao W, Powell J, Rosenwald A, Ott G, Muller-Hermelink HK, Gascoyne RD, Connors JM, Campo E, Jaffe ES, Delabie J, Smeland EB, Rimsza LM, Fisher RI, Weisenburger DD, Chan WC, Staudt LM (2008) Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc Natl Acad Sci USA* **105**: 13520–13525

Lockwood WW, Chari R, Coe BP, Girard L, Macaulay C, Lam S, Gazdar AF, Minna JD, Lam WL (2008) DNA amplification is a ubiquitous mechanism of oncogene activation in lung and other cancers. *Oncogene* **27**: 4615–4624

Luk C, Tsao MS, Bayani J, Shepherd F, Squire JA (2001) Molecular cytogenetic analysis of non-small cell lung carcinoma by spectral karyotyping and comparative genomic hybridization. *Cancer Genet Cytogenet* **125**: 87–99

Mead MN (2005) Arsenic: in search of an antidote to a global poison. *Environ Health Perspect* **113**: A378–A386

Moore LE, Smith AH, Eng C, Kalman D, DeVries S, Bhargava V, Chew K, Moore II D, Ferreccio C, Rey OA, Waldman FM (2002) Arsenic-related chromosomal alterations in bladder cancer. *J Natl Cancer Inst* **94**: 1688–1696

Nordstrom DK (2002) Public health. Worldwide occurrences of arsenic in ground water. *Science* **296**: 2143–2145

Nymark P, Wikman H, Ruosaari S, Hollmen J, Vanhala E, Karjalainen A, Anttila S, Knuutila S (2006) Identification of specific gene copy number changes in asbestos-related lung cancer. *Cancer Res* **66**: 5737–5743

Parsons JL, Preston BD, O'Connor TR, Dianov GL (2007) DNA polymerase delta-dependent repair of DNA single strand breaks containing 3′-end proximal lesions. *Nucleic Acids Res* **35**: 1054–1063

Pei J, Balsara BR, Li W, Litwin S, Gabrielson E, Feder M, Jen J, Testa JR (2001) Genomic imbalances in human lung adenocarcinomas and squamous cell carcinomas. *Genes Chromosomes Cancer* **31**: 282–287

Powell CA, Bueno R, Borczuk AC, Caracta CF, Richards WG, Sugarbaker DJ, Brody JS (2003) Patterns of allelic loss differ in lung adenocarcinomas of smokers and nonsmokers. *Lung Cancer* **39**: 23–29

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME (2006) Global variation in copy number in the human genome. *Nature* **444**: 444–454

Salnikow K, Zhitkovich A (2008) Genetic and epigenetic mechanisms in metal carcinogenesis and cocarcinogenesis: nickel, arsenic, and chromium. *Chem Res Toxicol* **21**: 28–44

Sambrook J, Fritsch E, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY

Samet JM, Avila-Tang E, Boffetta P, Hannan LM, Olivo-Marston S, Thun MJ, Rudin CM (2009) Lung cancer in never smokers: clinical epidemiology and environmental risk factors. *Clin Cancer Res* **15**: 5626–5645

Servicio de Salud Antofagasta (Unidad de Registro de Cancer) (2000) *III Anuario Registro Regional de Cancer Segunda Region Chile 2000 (Incluye información 1998–2000)*. Ministerio de Salud, Gobierno de Chile: Antofagasta, Chile

Smith AH, Goycolea M, Haque R, Biggs ML (1998) Marked increase in bladder and lung cancer mortality in a region of Northern Chile due to arsenic in drinking water. *Am J Epidemiol* **147**: 660–669

Sun S, Schiller JH, Gazdar AF (2007) Lung cancer in never smokers – a different disease. *Nat Rev Cancer* **7**: 778–790

Tonon G, Wong KK, Maulik G, Brennan C, Feng B, Zhang Y, Khatry DB, Protopopov A, You MJ, Aguirre AJ, Martin ES, Yang Z, Ji H, Chin L, Depinho RA (2005) High-resolution genomic profiles of human lung cancer. *Proc Natl Acad Sci USA* **102**: 9625–9630

Venkatesan RN, Treuting PM, Fuller ED, Goldsby RE, Norwood TH, Gooley TA, Ladiges WC, Preston BD, Loeb LA (2007) Mutation at the polymerase active site of mouse DNA polymerase delta increases genomic instability and accelerates tumorigenesis. *Mol Cell Biol* **27**: 7669–7682

Watson SK, deLeeuw RJ, Horsman DE, Squire JA, Lam WL (2007) Cytogenetically balanced translocations are associated with focal copy number alterations. *Hum Genet* **120**: 795–805

Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, Lam WL (2007) A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet* **80**: 91–104

Wong MP, Lam WK, Wang E, Chiu SW, Lam CL, Chung LP (2002) Primary adenocarcinomas of the lung in nonsmokers show a distinct pattern of allelic imbalance. *Cancer Res* **62**: 4464–4468

Yanagitani N, Kohno T, Kim JG, Kunitoh H, Tamura T, Takei Y, Tsuchiya S, Saito R, Yokota J (2003) Identification of D19S246 as a novel lung adenocarcinoma susceptibility locus by genome survey with 10-cM resolution microsatellite markers. *Cancer Epidemiol Biomarkers Prev* **12**: 366–371

Genetics and Genomics