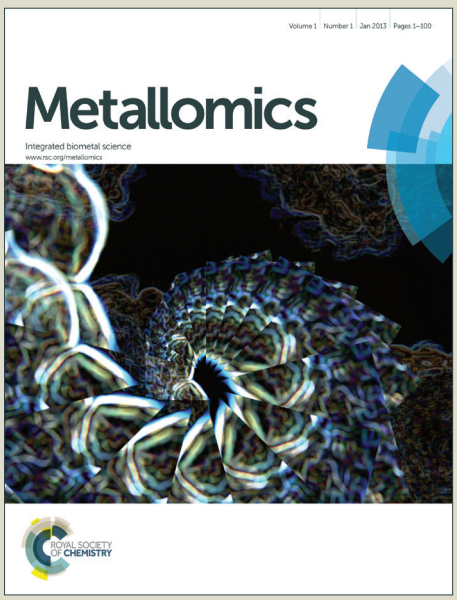
Metallomics

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/metallomics

Exposure to arsenic and intra-chromosomal instability in blood

Tomasz K Wojdacz [1], Matteo Bottai [2], Marie Vahter [1], and Karin Broberg [1]

[1] Unit of Metals and Health, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, SE – 171 77, Sweden

[2] Unit of Biostatistics, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, SE – 171 77, Sweden

Abstract:

The 450k Chip Analysis Methylation Pipeline (ChAMP) is a novel Illumina Infinium HumanMethylation450 BeadChip data processing algorithm that allows for analysis of copy number alterations (CNA). With this pipeline we evaluated the prevalence of CNA in peripheral blood leukocytes from healthy Argentinean Andean women with varying exposure to inorganic arsenic in drinking water. Arsenic exposure was assessed based on the sum concentrations of metabolites of inorganic arsenic in urine (U-As), which ranged 10-663µg/L with a median of 185 µg/L. We used linear regression analysis to elucidate the association between U-As and the prevalence of CNA. We found that increasing arsenic exposure was positively associated with the frequency of CNA ($p=0.002$), possibly in a dose-response relationship. Adjustment of the regression model for age and BMI of the subjects did not significantly change the effect estimate, although both covariates were significant predictors. Our results suggest that exposure to arsenic increases genomic instability in the form of CNA.

Genomic instability contributes to neoplastic transformation through deregulation of gene expression and activation of oncogenes (e.g. through chromosomal rearrangements) and inactivation of tumor suppressor genes (e.g. through deletions) [1-3]. It has been shown both *in vitro* and *in vivo* that exposure to inorganic arsenic induces genomic instability [4-10]. Bladder and lung tumors from patients exposed to arsenic were found to harbor higher levels of genomic instability than tumors from unexposed cancer patients, which hints at a causal link between arsenic induced genomic instability and carcinogenesis [2, 3, 11, 12]. Most studies of arsenic and genomic instability were performed at a cytogenetic level, and they showed an overrepresentation of numerical and structural chromosome aberrations as well as polyploidy with increasing arsenic exposure [5-8, 10]. Using a novel methodological approach, we sought to further the understanding of the genomic instability at the intra-chromosomal level in blood of subjects environmentally exposed to arsenic.

The 450k Chip Analysis Methylation Pipeline (ChAMP) is a novel Illumina Infinium HumanMethylation450 BeadChip data processing algorithm which allows for the evaluation of deferentially methylated regions between different groups of samples [13]. Moreover, the pipeline integrates a method for analysis of copy number alterations (CNA) on the basis of the beta values obtained for each CpG dinucleotide present on the 450k chip. In this analysis, Circular Binary Segmentation (CBS) algorithm from R-DNAcopy package was used to compute regions with differing copy number with a threshold derived from the difference between the log2 ratio between male and females of log2 +/-0.33 representing the change of one X chromosome [14]. The performance of the new approach in identifying CNAs was tested against Illumina CytoSNP chip, a SNP-based platform for identification of CNAs and the new pipeline enabled detection of 85% of the CNAs [14]. We applied this

algorithm in our study to assess the influence of environmental exposure on the prevalence of CNAs in blood.

Our study involved women living in area with varying arsenic concentrations in drinking water in the Andean part of Salta Province in Northern Argentina. Detailed description of the study area [15] and the study population [16] were published previously. A group of 172 nonrelated women were recruited in 2008. Exposure to arsenic was assessed based on the concentrations of metabolites of inorganic arsenic (methylarsonic acid (%MMAs); dimethylarsinic acid (%DMAs); and remaining inorganic arsenic (%iAs) in urine, as measured by high pressure liquid chromatography, on line with hydride generation and inductively coupled plasma mass spectrometry (ICP-MS) [17, 18]. For the 450K chip analysis, we selected 91 of the women with a mean U-As of 185 µg/L (10-663.8µg/L) and median 185 µg/L. The women had a similar diet and ethnic background. They did not drink alcohol and only three reported smoking (moderately).

We first plotted the CNAs against U-As and found an increase in CNA with increasing U-As, with an apparent linear relationship (Figure 1). We therefore used linear regression analysis to model the association between U-As and the prevalence of CNAs. The regression analyses showed an increase in the number of CNAs with increasing U-As (p=0.002, Table 1). The estimate indicated one extra CNA for 100 µg/L of U-As.. Also, adjustment for age and BMI did not markedly change the effect estimate. We also modeled the influence of arsenic metabolism efficiency by including fractions of metabolites in the models but this did not significantly influence the association between U-As and CNAs. It is important to mention here that we previously showed that subjects in this study are likely to have developed genetic adaptation to high-arsenic living conditions[19]. Consequently the effect

observed in this study may be smaller than the effect observed in other populations, but further studies are needed to elaborate this hypothesis.

Secondly, we analyzed the length and type of the CNAs present in blood for each subject. The median length of the CNA in the samples was 134 bp (ranging from 2 to 590332 bp) which corresponds to the size of microsatellite repeats, short deletions/amplifications to copy number variation (>1000 bp). At the same time over 60% of total 2410 CNAs detected across all the samples were deletions indicating that arsenic exposure is more likely to induce this type of alternations (for details see Supplementary data 1). We also used coordinates for each CNA in a given sample (a part of the ChAMP output) and mapped all significant CNAs regions to genome with Integrated Genomic Viewer [20]. The CNA regions were located on all chromosomes, but there was an overrepresentation of CNAs on chromosomes 6 and 17 considering the chromosome length. Most of the CNAs at chromosome 6 were located in p22.1-p21.32 region, which includes the highly variable HLA loci. The CNAs on chromosome 17 did not involve the region harboring *TP53* encoding the p53 protein that has been shown to be affected by arsenic exposure[21]. We also computed the most common regions displaying CNAs across our samples and found that 496 common CNA regions (ranging from 2 to 17942bp) were present in more than 2 samples and 32 of those CNAs were present in more than 20 samples. The principles of computing of most common regions affected by CNA among individual samples are depicted in Figure 2. Detailed description and coordinates of loci harboring those alterations are shown in Supplementary data 1. Functional analyses of those specific loci are ongoing and out of scope of this short communication.

Conclusions

This is the first study that utilizes the novel Infinium HumanMethylation450 BeadChip data processing algorithm for investigation of CNA in relation to an environmental exposure. The main limitation of our analysis is that Illumina Infinium HumanMethylation450 BeadChip was designed to cover the regions of the genome with high frequency of CpG sites. Consequently, CpG poor regions, which constitute a vast part of the genome, are not interrogated in our analysis. This limitation, however, can be overcome in the future by using SNP-based arrays, considering the whole genome. We anticipate that such analyses will significantly strengthen our initial findings.

The findings presented here corroborate previous cytogenetic findings and for the first time indicate that chronic exposure to arsenic may promote intra-chromosomal instability. The actual mechanism of this phenomenon needs further investigation. However, it can be speculated that arsenic indirectly contribute to this type of instability by generation of free radicals which, in turn, damage the DNA [18, 22, 23]. Also, arsenic may inhibit the DNA repair machinery, nucleotide excision repair (NER) [24], and/or base excision repair (BER) [25], which, if repressed, could result in intra-chromosomal rearrangements (as reviewed in: [9])

Overall, our results show that the ChAMP data processing pipeline can be a robust and economic tool to identify genomic instability in a biologically relevant context.

Figure legends

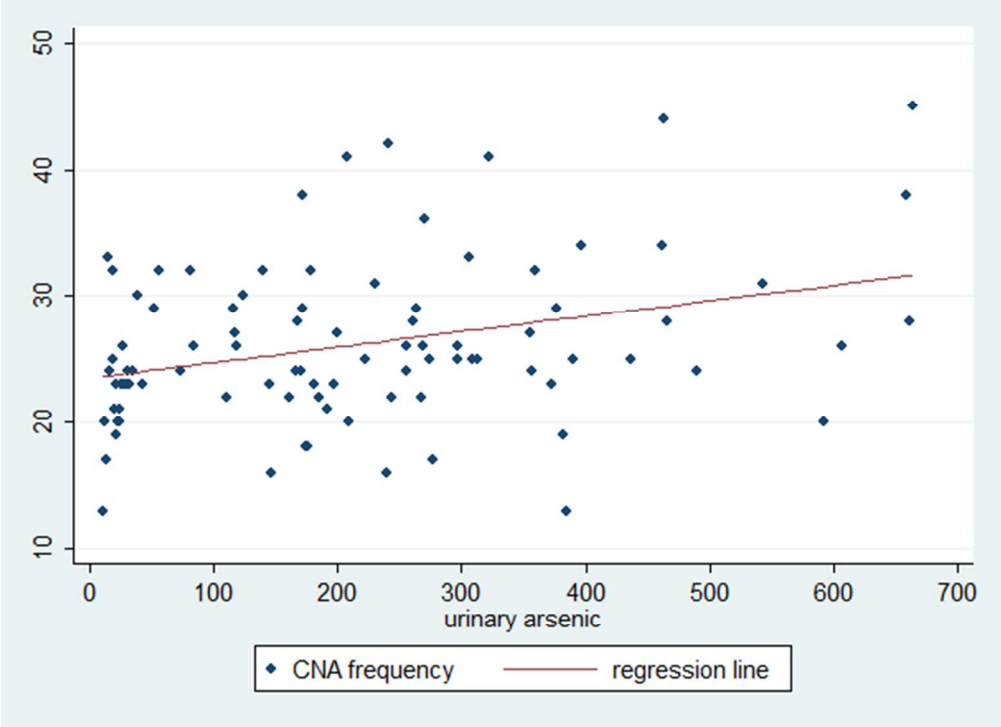
Figure 1. Scatter plot of CNAs frequency in peripheral blood versus U-As in Andean women with regression line

Figure 2. Illustration of the selection procedure for common CNAs across all the samples

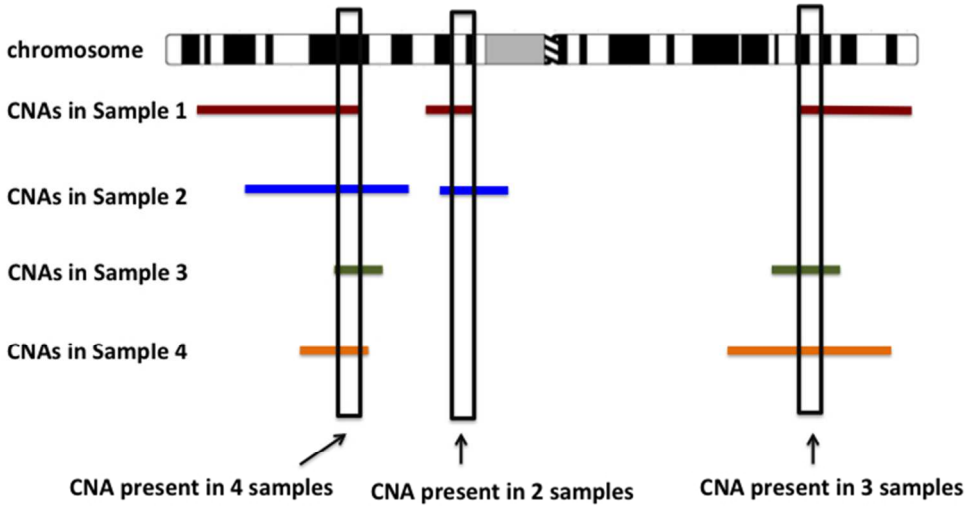
References

1. M. Fanciulli, E. Petretto and T. J. Aitman, *Clinical genetics*, 2010, 77, 201-213.
2. D. Hanahan and R. A. Weinberg, *Cell*, 2011, 144, 646-674.
3. F. Mitelman, B. Johansson and F. Mertens, *Nature reviews. Cancer*, 2007, 7, 233-245.
4. C. M. Steinmaus, C. Ferreccio, J. A. Romo, Y. Yuan, S. Cortes, G. Marshall, L. E. Moore, J. R. Balmes, J. Liaw, T. Golden and A. H. Smith, *Cancer Epidemiol Biomarkers Prev*, 2013, 22, 623-630.
5. S. Ghosh, G. Kewalramani, G. Yuen, T. Pulinilkunnil, D. An, S. M. Innis, M. F. Allard, R. B. Wambolt, D. Qi, A. Abrahani and B. Rodrigues, *Free radical biology & medicine*, 2006, 41, 1413-1424.
6. M. E. Gonsebatt, L. Vega, A. M. Salazar, R. Montero, P. Guzman, J. Blas, L. M. Del Razo, G. Garcia-Vargas, A. Albores, M. E. Cebrian, M. Kelsh and P. Ostrosky-Wegman, *Mutation research*, 1997, 386, 219-228.
7. G. Sciandrello, R. Barbaro, F. Caradonna and G. Barbata, *Mutagenesis*, 2002, 17, 99-103.
8. G. Sciandrello, F. Caradonna, M. Mauro and G. Barbata, *Carcinogenesis*, 2004, 25, 413-417.
9. P. Bhattacharjee, M. Banerjee and A. K. Giri, *Environment international*, 2013, 53, 29-40.
10. F. N. Dulout, C. A. Grillo, A. I. Seoane, C. R. Maderna, R. Nilsson, M. Vahter, F. Darroudi and A. T. Natarajan, *Mutat Res*, 1996, 370, 151-158.
11. L. E. Moore, A. H. Smith, C. Eng, D. Kalman, S. DeVries, V. Bhargava, K. Chew, D. Moore, 2nd, C. Ferreccio, O. A. Rey and F. M. Waldman, *Journal of the National Cancer Institute*, 2002, 94, 1688-1696.
12. R. Hubaux, D. D. Becker-Santos, K. S. Enfield, D. Rowbotham, S. Lam, W. L. Lam and V. D. Martinez, *Molecular cancer*, 2013, 12, 20.
13. T. J. Morris, L. M. Butcher, A. Feber, A. E. Teschendorff, A. R. Chakravarthy, T. K. Wojdacz and S. Beck, *Bioinformatics*, 2014, 30, 428-430.
14. A. Feber, P. Guilhamon, M. Lechner, T. Fenton, G. A. Wilson, C. Thirlwell, T. J. Morris, A. M. Flanagan, A. E. Teschendorff, J. D. Kelly and S. Beck, *Genome biology*, 2014, 15, R30.
15. G. Concha, K. Broberg, M. Grander, A. Cardozo, B. Palm and M. Vahter, *Environmental science & technology*, 2010, 44, 6875-6880.
16. K. Engstrom, M. Vahter, S. J. Mlakar, G. Concha, B. Nermell, R. Raqib, A. Cardozo and K. Broberg, *Environmental health perspectives*, 2011, 119, 182-188.
17. B. Fangstrom, J. Hamadani, B. Nermell, M. Grander, B. Palm and M. Vahter, *Toxicol Appl Pharmacol*, 2009, 239, 208-214.
18. K. S. Engstrom, M. Vahter, C. Lindh, F. Teichert, R. Singh, G. Concha, B. Nermell, P. B. Farmer, U. Stromberg and K. Broberg, *Mutat Res*, 2010, 683, 98-105.
19. C. M. Schlebusch, C. M. Lewis, Jr., M. Vahter, K. Engstrom, R. Y. Tito, A. J. Obregon-Tito, D. Huerta, S. I. Polo, A. C. Medina, T. D. Brutsaert,

G. Concha, M. Jakobsson and K. Broberg, *Environmental health perspectives*, 2013, 121, 53-58.
20. H. Thorvaldsdottir, J. T. Robinson and J. P. Mesirov, *Briefings in bioinformatics*, 2013, 14, 178-192.
21. T. G. Rossman and C. B. Klein, *Metallomics : integrated biometal science*, 2011, 3, 1135-1141.
22. S. J. Flora, *Free radical biology & medicine*, 2011, 51, 257-281.
23. T. G. Rossman, *Mutation research*, 2003, 533, 37-65.
24. A. S. Andrew, J. L. Burgess, M. M. Meza, E. Demidenko, M. G. Waugh, J. W. Hamilton and M. R. Karagas, *Environmental health perspectives*, 2006, 114, 1193-1198.
25. F. Ebert, A. Weiss, M. Bultemeyer, I. Hamann, A. Hartwig and T. Schwerdtle, *Mutation research*, 2011, 715, 32-41.



213x155mm (72 x 72 DPI)



254x190mm (72 x 72 DPI)

Table 1. Description of A) variables and B) association between urinary arsenic concentrations (U-As) and average frequency of CNAs

A	**Variable description mean (range)**		
U-As	226 (10-1251) µg/L		
CNAs	26 (13-45)		
Age	32 (12-64) years		
BMI	23.8 (16.4-35.2) kg/m		
B	Model	β_1 (95% CI)	P value
U-As	Average CNA-frequency = $\alpha + \beta_1$ (U-As)	0.010 (0.003; 0.016)	0.002
Age/BMI	Average CNA-frequency = $\alpha + \beta_1$ (U-As) + β_2 age + β_3 BMI	0.010 (0.004; 0.016)	0.001