Title: Multiple aberrations of chromosome 3p detected in oral premalignant lesions

Authors: Ivy F. L. Tsui[1, 3], Miriam P. Rosin[2], Lewei Zhang[4], Raymond T. Ng[5], Wan L. Lam[1, 3]

Affiliations: Departments of [1]Cancer Genetics and Developmental Biology, and [2]Cancer Control Research, British Columbia Cancer Research Centre, Vancouver, British Columbia, Canada; Departments of [3]Pathology and Laboratory Medicine, [4]Oral Biological and Medical Sciences, and [5]Computer Science, University of British Columbia, Vancouver, British Columbia, Canada

Running title: Segmental changes on the 3p arm in oral premalignant lesions

Keywords: oral premalignant lesions, chromosome 3p, array CGH

Requests for reprints: Ivy Tsui, BC Cancer Research Centre, 675 West 10th Avenue, Vancouver, BC, V5Z 1L3, Canada; Phone: 604-675-8111; Fax: 604-675-8283; Email: itsui@bccrc.ca

Abstract

The study of oral premalignant lesions (OPLs) is crucial to the identification of initiating genetic events in oral cancer. However, these lesions are minute in size, making it a challenge to recover sufficient DNA from microdissected cells for comprehensive genomic analysis. As a step toward identifying genetic aberrations associated with oral cancer progression, we used tiling-path array comparative genomic hybridization (CGH) to compare alterations on chromosome 3p for 71 OPLs against 23 oral squamous cell carcinomas (OSCCs). 3p was chosen because, while it is frequently altered in oral cancers and has been associated with progression risk, its alteration status has only been evaluated at a small number of loci in OPLs. We identified six recurrent losses in this region that were shared between high-grade dysplasias (HGDs) and OSCCs, including a 2.89 Mbp deletion spanning the *FHIT* gene (previously implicated in oral cancer progression). When the alteration status for these six regions was examined in 24 low-grade dysplasias (LGDs) with known progression outcome, we observed that they occurred at a significantly higher frequency in LGDs that later progressed to later stage disease ($p < 0.003$). Moreover, parallel analysis of all profiled tissues showed that the extent of overall genomic alteration at 3p increased with histological stage. This first high resolution analysis of chromosome arm 3p in oral premalignant lesions represents a significant step towards predicting progression risk in early pre-invasive disease and provides a keen example of how genomic instability escalates with progression to invasive cancer.

Introduction

Loss of chromosome 3p is a common genetic event in many human cancers, with several putative tumor suppressor genes (TSGs) located in this region (1, 2). In oral cancer, loss of 3p carries prognostic significance, particularly for risk of disease progression, development of second primary tumors, and emergence of local recurrences (3-7). Alterations on 3p are also thought to be key events in the progression of oral premalignant lesions (OPLs) to invasive disease. Previously, three microsatellite markers on this chromosome arm were used to demonstrate that 3p loss of heterozygosity (LOH) did in fact occur in OPLs and that alteration of this region increased in frequency with progression to cancer (8). Other reports have also associated alterations in specific segments of 3p in OPLs with elevated risk of invasive transformation (4, 6). However, these markers have not been applicable in all cases, leaving the possibility that there are additional candidates within chromosome 3p that are necessary for progression to oral invasive cancer.

To identify additional segments on chromosome 3p that are associated with disease progression, we undertook tiling-path array comparative genomic hybridization (CGH) for the entire chromosome arm in a panel of premalignant and invasive oral tissues. In addition to its improved resolution, this platform also gave the benefit of being able to work with DNA of limited quantity (because of the small size of the captured lesions) and low quality (because of the use of formalin-fixed paraffin-embedded tissue, which typically precludes effective microarray profiling). This allowed us to profile high-grade dysplasias (HGDs) (including severe dysplasias and carcinoma *in situ* [CIS] lesions, which have a high likelihood of progression to invasive disease), low-grade dysplasias (LGDs) with clinical outcome, and oral squamous cell carcinomas (OSCCs). Parallel analysis of genomic data from these various tissues

demonstrated stage-specific genetic alterations. More importantly, association of these same data with clinical features provided us with a new tool for predicting progression risks in LGDs, where the existing histopathological criteria are unable to predict progression.

Materials and Methods

Tissue Samples. This study involves 94 archival FFPE specimens from 86 patients: 24 LGDs (low-grade dysplasias: 2 hyperplasias, 22 mild and moderate dysplasias), 47 HGDs (severe dysplasias and *CIS*), and 23 OSCCs, all obtained from the British Columbia Oral Biopsy Service. Clinical information and demographics for these cases are presented in Supplementary Table S1. All LGDs and HGDs came from patients with no prior history of cancer. All LGD patients were enrolled in a longitudinal study established at the British Columbia Cancer Prevention Program with a median follow up of seven years: of the 24 LGDs, 15 LGDs did not progress during the period from 1985 to 2007, whereas nine LGDs progressed to cancer. In British Columbia patients with HGDs are treated with surgery and followed for recurrence. All diagnoses were confirmed by the study pathologist (L. Z.). Representative sections were micro-dissected and DNA was extracted as previously described (9).

Array CGH analysis. Array CGH was performed as previously described (9-11). Genomic arrays (SMRT v.1 and v.2) were obtained from the BC Cancer Research Center Array Laboratory (9, 10). Briefly, sample and normal reference genomic DNA (250 ng each) were differentially labeled and mixed with 100 µg of human Cot-1 DNA (Invitrogen, Ontario), purified and hybridized to the array at 45ºC for 36 hours before washing. Hybridized arrays were scanned for signal intensities as previously described (9, 12). Image data were LOWESSS, spatial, and median normalized (13, 14). *SeeGH* software combined duplicate spot data and

displayed $\log_2$ ratios in relation to genomic locations in the hg17 assembly (NCBI Build 35)

(15). Duplicate spots with standard deviations >0.075 or with signal-to-noise ratios <3 were

stringently filtered, leaving informative clones for breakpoint analysis (9, 12). Segmental gains

and losses were detected and confirmed using three separate algorithms: *aCGH-Smooth*,

DNACopy, and *LSPHMM* (16-18). Altered regions that were concurrently identified by at least

two of the three segmentation algorithms were scored: (i) *aCGH-Smooth* uses a local search

algorithm that smoothes the observed array CGH values between consecutive breakpoints to a

suitable common value by using maximum likelihood estimation (16). The Lambda and the

maximum number of breakpoints in initial pool were set to 6.75 and 100 respectively (12). (ii)

DNAcopy employs a circular binary segmentation algorithm that uses a random permutation test

to determine the statistical significance of change points (17). Default settings were used to

delineate sections of differing copy number (13, 17). (iii) *LSPHMM (location-specific priors*

hidden Markov models) uses a modification of hidden Markov models to exploit prior

knowledge about location of copy number polymorphisms which were often detected as outliers

in standard HMMs (18). Altered regions were further confirmed by visual inspection. To avoid

false-positives due to hybridization "noise", only those alterations containing ≥3 overlapping

clones were considered as segmental loss. Whole arm loss was defined as a loss that

encompasses all BAC clones from the telomeric to the centromeric end of the 3p arm. A

summary of filtering criteria and genetic pattern detected for each sample is listed in

Supplementary Table S2. Filtering threshold was the same for all except for two samples

(Supplementary Table S2).

Copy number variation (CNV). Detected alterations were filtered for CNVs that had been

previously identified using the same array platform (11). The list of CNV-associated BAC

clones has been made publicly available on the UCSC Genome Browser and has also been incorporated into *LSPHMM* software (11, 18). Examples of CNVs in normal human individuals are presented in Supplementary Fig. S1.

Statistical analysis. Proportion-test was used to test if the number of samples that show 3p losses in HGDs and OSCCs were statistically distinguishable. Wilcoxon rank-sum test was employed to calculate the probability distributions of the proportion of altered regions between each sample group were different due to chance alone. A two-tailed Fisher's exact test was applied to calculate the probability that the observed association for the number of genetic alterations between non-progressing and progressing LGDs, LGDs and HGDs, and HGDs and OSCCs has occurred by chance alone. The significance threshold was set at $p < 0.05$. All statistical tests were two-sided and were performed in the R statistical environment (v2.6.2).

Loss of heterozygosity (LOH) analysis. LOH analysis focused on a region previously shown to be associated with increased risk of progression, utilizing microsatellite markers *D3S1234*, *D3S1228*, and *D3S1300* which are mapped to 3p14.2-3p14.1 (6). Microsatellite analyses were performed as previously described and LOH was inferred when the signal ratio of the two alleles differed in the normal samples by at least 50% (6). Microsatellite markers that yielded homozygous alleles were termed "non-informative".

Results and Discussion

We hypothesized that genetic alterations critical for tumorigenesis might be found at the premalignant stage of oral lesions. To test this, we analyzed the 3p arm of 71 OPLs at high resolution to discover changes that occur during the progression of OPLs. Six discrete, recurrently altered regions were identified in 47 HGDs. These regions were then compared

against 24 LGDs and 23 OSCCs to determine the prevalence of such changes in different stages

of progression. The set of 94 profiles have been deposited to GEO database at NCBI, series

accession number GSE9193 (reviewer's link:

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=rzinvqmkacueujy&acc=GSE9193).

Graphical representations of the data are available at www.bccrc.ca/cg/ArrayCGH_Group.html.

Segmental alterations on chromosome 3p are more frequent than whole arm loss in HGDs.

We employed three independent breakpoint-detection algorithms on each sample to identify

common regions of loss. Only regions identified by 2/3 algorithms and confirmed by visual

inspection were considered to be lost. In addition, because of the existence of CNVs in the

normal human population, previously identified clones associated with copy number

polymorphisms were filtered.

Our data showed that small regions of 3p were commonly lost in HGDs. Regions of

segmental genetic loss were found in 80.8% of HGDs, with whole arm loss apparent in only a

quarter of cases (12/47) and segmental alterations in 55.3% (26/47 cases). Segmental alterations

showed great variability in size, ranging from 5.3 to 88.7 Mbp (Fig. 1*A*). Genetic gains were

infrequent, with only two HGDs showing such changes. The number of 3p alterations in HGDs

(38/47) and OSCCs (23/23) was not statistically different (difference in proportions: -0.15, 95%

confidence interval -0.32 to 0.02, $p = 0.19$, proportion-test), with this evidence of similarly

increased genetic instability perhaps accounting for the relatively high progression risk for

HGDs. The most striking difference in gene alterations between HGDs and OSCCs was the

alteration size for each group. While loss of the whole 3p arm is a common occurrence in

OSCCs (3, 9), alterations in HGD lesions did not typically encompass the entire chromosome

arm. The size of 3p alterations in OSCCs and HGDs was significantly different ($p = 1.84 \times 10^{-5}$, Wilcoxon rank-sum test).

Regions of loss in HGDs and comparison to OSCCs. Those 3p alterations apparent in both HGDs and OSCCs are more likely to represent key genomic changes underlying early oral tumorigenesis. We identified six minimally altered regions (MARs) that were recurrent in at least 75% of HGD cases that harbored segmental alterations (Fig. 1*A*). These occurred at 3p25.3-p26.1, 3p25.1-p25.3, 3p24.1, 3p21.31-p22.3, 3p14.2, and 3p14.1 (regions A to F in Fig. 1*A*, Table 1). The data of the upper and lower boundary of each MAR is illustrated in Supplementary Fig. S2. Most identified MARs harbored <10 genes. Significantly, alteration of the previously described oral TSG *FHIT* (19) occurred at the same high frequency as the other identified MARs, strongly suggesting the presence of additional candidates driving oral tumorigenesis (Table 1, Fig. 1*C*). Analysis of all OSCCs revealed loss of the entire 3p arm in 19/23 cases, a frequency consistent with previous reports (9, 20-25). The presence of expanded deletions between these stages indicates the occurrence of secondary genetic events from OPLs to invasive disease. This also shows the value of analyzing premalignant lesions in order to identify early genetic events that are likely masked by subsequent accumulated alterations.

Evidence of segmental alterations in progressing LGDs. We next examined progressing LGDs (9/24, median follow-up time for all cases >7 years) to determine if they accumulate more and/or different genetic alterations than their morphologically similar non-progressing counterparts (15/24). Segmental losses were significantly associated with progression, occurring in 7/9 progressing LGDs and only 2/15 non-progressing LGDs ($p = 0.003$, Fisher's exact test).

These findings are consistent with previous reports of higher frequency of LOH at 3p microsatellite loci in progressing OPLs (5-7). We also observed that the size of segmental losses were larger for progressing LGD cases (53.7 Mbp versus 0.78 Mbp), with the distribution of losses among non-progressing (n = 15) and progressing LGDs (n = 9) being significantly different ($p = 6.1 \times 10^{-5}$, Wilcoxon rank-sum test). Interestingly, the aberrations in progressing LGDs resemble those of HGDs (which, as already stated, carry a higher likelihood of progression to invasive disease) (Fig. 1*B*). To examine this further, we evaluated the LGD cases for the alteration status of the six MARs identified in HGDs. As anticipated, these regions were found to be frequently altered in progressing LGDs but not in non-progressing LGDs ($p < 0.003$, Fig. 1*B*). These data lend credence to the conclusion that the MARs identified in HGDs represent critical events in progression to invasive disease.

Disrupted genes in dysplastic lesions. One hundred forty-one annotated genes were spanned by the six MARs identified above (RefSeq release 25, see listing in Supplementary Table S3). Twenty-eight of these are known to be associated with cancer and 17 have functions potentially related to tumor suppressor activity (Table 1). Specifically, nine of these genes have actually been reported as TSGs in cancer (*TIMP4*, *PPARG*, *WNT7A*, *TGFBR2*, *MLH1*, *CTDSPL*, *VILL*, *AXUD1*, and *FHIT*) (26-34). The 17 candidate TSGs can be generally categorized to have functions involved in the inhibition of cAMP signaling for the induction of apoptosis (*GRM7*), DNA-damage repair (*RAD18*, *XPC*), autophagy-related and ubiquitin-activating enzyme activity (*ATG7*), metalloendopeptidase inhibitor activity (*TIMP4*), transcription repression activity (*PPARG)*, apoptosis (*ACVR2B*, *AXUD1*), MAPK signal transduction (*SPGAP3*, *TGFBR2*), Wnt pathway-mediated anti-tumorigenic signaling (*LRRFIP2*, *WNT7A*, *ACVR2b*, and *AXUD1*), and

negative regulation of progression through cell cycle (*MLH1*, *DLEC1*, *FHIT*). Specifically, *GRM7* was shown to be methylated in chronic lymphocytic leukemia (35), whereas *TGFBR2* is transcriptionally-repressed in human epithelial tumors (36).

Our candidate MARs are in agreement with previous studies of OPLs. Regions A, B, D, E, and F coincide with reported allelic losses identified by microsatellite analysis or copy number loss by conventional CGH (5, 6, 8, 22, 24, 25). The improved resolution in the current work has further fine-mapped these regions and allows us to name specific gene candidates (25). The frequent deletion of 3p24.1 has been delineated to 1.21 Mbp, highlighting the potential importance of *TGFBRII* in oral tumorigenesis (Table 1).

Sequential 3p deletions during OPL development. We examined the 3p arm for frequency of alteration, genetic pattern, and size of change during the multistep oral tumorigenesis. The increased frequency of 3p loss from LGDs (37.5%) to HGDs (80.8%) was significant ($p = 4.6$ x 10^{-4}, Fisher's exact test). Also, specific genetic patterns were associated with each histological group (Fig. 2). Whole chromosome arm changes were observed at much higher frequency for OSCCs tumors (Table 2), with the mean deletion size increasing significantly with disease stage; the average deletion was 41.9 Mbp in LGDs, 67.5 Mbp in HGDs, and 83.5 Mbp in OSCCs ($p = 0.016$ for LGDs versus HGDs and $p = 3.6$ x 10^{-4} for HGDs versus OSCCs, Wilcoxon rank-sum test). Genetic changes on 3p are sequentially associated with increased progressing stages, as most LGDs have no loss, HGDs have segmental losses, and OSCCs have whole arm loss. These data clearly delineate escalation of genomic instability at chromosome 3p with histopathological disease stage. (No associations between age or gender were observed for our genomic data, and

likewise, smoking status was not found to influence the level of genetic instability on the 3p arm in our dataset (Supplementary Table S4)).

Integrating LOH data with copy number alterations. DNA copy number loss can be detected by array CGH, while segmental loss or uniparental disomy can be identified by LOH analyses. To explore the relative contributions of each kind of genetic dysregulation to OPLs and OSCCs, we integrated copy number data with informative LOH data for 3p14.1-p14.2 (available for 88/94 case). Concordant LOH and copy number loss were observed for 32.9% of these cases, with a further 30.7% exhibiting copy number neutral LOH and 12.5% showing copy number loss not detected by LOH (Supplementary Fig. S3). Complementary integrative analysis by these two approaches facilitates a more complete portrait of genome dysregulation in oral cancer progression and could represent an even more powerful tool for delineating progression risk in oral premalignant lesions.

Summary. Alteration of chromosome arm 3p has previously been reported as a frequent event for oral cancer and losses of specific loci within this region have been associated with progression risk in early OPLs. To date, however, there have been no high resolution analyses of this chromosome arm to identify additional candidates that may drive oral tumorigenesis. By undertaking tiling-path array CGH analysis for a panel of nearly a hundred OPLs and oral tumors, we have confirmed that genetic instability increases on this single arm with progression to invasive cancer, identified new candidate TSGs that may contribute to oral cancer progression, and demonstrated that low grade dysplasias known to progress to invasive disease harbor gene alterations very similar to those observed in higher grade dysplastic lesions. This study

identified specific regions altered on chromosome 3p that are commonly found in the premalignant stages of oral cancer development but are often masked in invasive tumors due to increased genetic instability.

Acknowledgments

References

1. Garnis C, Buys TP,Lam WL. Genetic alteration and gene expression modulation during cancer progression. Mol Cancer 2004;3:9.
2. Zabarovsky ER, Lerman MI,Minna JD. Tumor suppressor genes on chromosome 3p involved in the pathogenesis of lung and other cancers. Oncogene 2002;21:6915-35.
3. Mao L, Hong WK,Papadimitrakopoulou VA. Focus on head and neck cancer. Cancer Cell 2004;5:311-6.
4. Mao L, Lee JS, Fan YH, et al. Frequent microsatellite alterations at chromosomes 9p21 and 3p14 in oral premalignant lesions and their value in cancer risk assessment. Nat Med 1996;2:682-5.
5. Partridge M, Pateromichelakis S, Phillips E, Emilion GG, A'Hern RP,Langdon JD. A case-control study confirms that microsatellite assay can identify patients at risk of developing oral squamous cell carcinoma within a field of cancerization. Cancer Res 2000;60:3893-8.
6. Rosin MP, Cheng X, Poh C, et al. Use of allelic loss to predict malignant risk for low-grade oral epithelial dysplasia. Clin Cancer Res 2000;6:357-62.
7. Rosin MP, Lam WL, Poh C, et al. 3p14 and 9p21 loss is a simple tool for predicting second oral malignancy at previously treated oral cancer sites. Cancer Res 2002;62:6447-50.
8. Califano J, van der Riet P, Westra W, et al. Genetic progression model for head and neck cancer: implications for field cancerization. Cancer Res 1996;56:2488-92.
9. Baldwin C, Garnis C, Zhang L, Rosin MP,Lam WL. Multiple microalterations detected at high frequency in oral cancer. Cancer Res 2005;65:7561-7.
10. Ishkanian AS, Malloff CA, Watson SK, et al. A tiling resolution DNA microarray with complete coverage of the human genome. Nat Genet 2004;36:299-303.
11. Wong KK, deLeeuw RJ, Dosanjh NS, et al. A comprehensive analysis of common copy-number variations in the human genome. Am J Hum Genet 2007;80:91-104.
12. Coe BP, Lockwood WW, Girard L, et al. Differential disruption of cell cycle pathways in small cell and non-small cell lung cancer. Br J Cancer 2006;94:1927-35.
13. Chari R, Lockwood WW,Lam WL. Computational Methods for the Analysis of Array Comparative Genomic Hybridization. Cancer Inform 2006;2:48-58.
14. Khojasteh M, Lam WL, Ward RK,MacAulay C. A stepwise framework for the normalization of array CGH data. BMC Bioinformatics 2005;6:274.
15. Chi B, DeLeeuw RJ, Coe BP, MacAulay C,Lam WL. SeeGH--a software tool for visualization of whole genome array comparative genomic hybridization data. BMC Bioinformatics 2004;5:13.
16. Jong K, Marchiori E, Meijer G, Vaart AV,Ylstra B. Breakpoint identification and smoothing of array comparative genomic hybridization data. Bioinformatics 2004;20:3636-7.
17. Olshen AB, Venkatraman ES, Lucito R,Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 2004;5:557-72.
18. Shah SP, Xuan X, DeLeeuw RJ, et al. Integrating copy number polymorphisms into array CGH analysis using a robust HMM. Bioinformatics 2006;22:e431-9.
19. Lee JI, Soria JC, Hassan K, et al. Loss of Fhit expression is a predictor of poor outcome in tongue cancer. Cancer Res 2001;61:837-41.

20.   Smeets SJ, Braakhuis BJ, Abbas S, et al. Genome-wide DNA copy number alterations in head and neck squamous cell carcinomas with or without oncogene-expressing human papillomavirus. Oncogene 2006;25:2558-64.

21.   Snijders AM, Schmidt BL, Fridlyand J, et al. Rare amplicons implicate frequent deregulation of cell fate specification pathways in oral squamous cell carcinoma. Oncogene 2005;24:4232-42.

22.   Chakraborty SB, Dasgupta S, Roy A, et al. Differential deletions in 3p are associated with the development of head and neck squamous cell carcinoma in Indian patients. Cancer Genet Cytogenet 2003;146:130-8.

23.   Garnis C, Baldwin C, Zhang L, Rosin MP,Lam WL. Use of complete coverage array comparative genomic hybridization to define copy number alterations on chromosome 3p in oral squamous cell carcinomas. Cancer Res 2003;63:8582-5.

24.   Kayahara H, Yamagata H, Tanioka H, Miki T,Hamakawa H. Frequent loss of heterozygosity at 3p25-p26 is associated with invasive oral squamous cell carcinoma. J Hum Genet 2001;46:335-41.

25.   Noutomi Y, Oga A, Uchida K, et al. Comparative genomic hybridization reveals genetic progression of oral squamous cell carcinoma from dysplasia via two different tumourigenic pathways. J Pathol 2006;210:67-74.

26.   Riddick AC, Shukla CJ, Pennington CJ, et al. Identification of degradome components associated with prostate cancer progression by expression analysis of human prostatic tissues. Br J Cancer 2005;92:2171-80.

27.   Haider AS, Peters SB, Kaporis H, et al. Genomic analysis defines a cancer-specific gene expression signature for human squamous cell carcinoma and distinguishes malignant hyperproliferation from benign hyperplasia. J Invest Dermatol 2006;126:869-81.

28.   Ishiguro H, Tsunoda T, Tanaka T, Fujii Y, Nakamura Y,Furukawa Y. Identification of AXUD1, a novel human gene induced by AXIN1 and its reduced expression in human carcinomas of the lung, liver, colon and kidney. Oncogene 2001;20:5062-6.

29.   Kashuba VI, Li J, Wang F, et al. RBSP3 (HYA22) is a tumor suppressor gene implicated in major epithelial malignancies. Proc Natl Acad Sci U S A 2004;101:4906-11.

30.   Kujan O, Oliver R, Roz L, et al. Fragile histidine triad expression in oral squamous cell carcinoma and precursor lesions. Clin Cancer Res 2006;12:6723-9.

31.   Michalik L, Desvergne B,Wahli W. Peroxisome-proliferator-activated receptors and cancers: complex stories. Nat Rev Cancer 2004;4:61-70.

32.   Sengupta S, Chakrabarti S, Roy A, Panda CK,Roychoudhury S. Inactivation of human mutL homolog 1 and mutS homolog 2 genes in head and neck squamous cell carcinoma tumors and leukoplakia samples by promoter hypermethylation and its relation with microsatellite instability phenotype. Cancer 2007;109:703-12.

33.   Winn RA, Van Scoyk M, Hammond M, et al. Antitumorigenic effect of Wnt 7a and Fzd 9 in non-small cell lung cancer cells is mediated through ERK-5-dependent activation of peroxisome proliferator-activated receptor gamma. J Biol Chem 2006;281:26943-50.

34.   You KT, Li LS, Kim NG, et al. Selective translational repression of truncated proteins from frameshift mutation-derived mRNAs in tumors. PLoS Biol 2007;5:e109.

35.   Rush LJ, Raval A, Funchain P, et al. Epigenetic profiling in chronic lymphocytic leukemia reveals novel methylation targets. Cancer Res 2004;64:2424-33.

36.   Bierie B,Moses HL. Tumour microenvironment: TGFbeta: the molecular Jekyll and Hyde of cancer. Nat Rev Cancer 2006;6:506-20.

Figure Legends

Figure 1. Frequency of alterations on the 3p arm. *A*, summary of copy number changes in 26 HGDs that showed segmental alterations on 3p. Copy number loss is presented as vertical lines on the left side of the ideogram, with vertical lines on the right side indicating copy number gain. For each sample, altered regions that were concurrently identified by at least two of the three methods were scored and confirmed by visual inspection. The top six minimal altered regions (MARs) are boxed in orange. *B*, frequency of alteration of the six MARs in 23 OSCCs (black), 47 HGDs (purple), and 24 LGDs (progressing LGDs in light purple and non-progressing LGDs in white). The six MARs were found to be common in progressing LGDs but infrequent in non-progressing LGDs ($p < 0.003$, Fisher's exact test). *C*, Alignment of four HGDs reveals segmental loss of the *FHIT* locus at 3p14.2. Each bar represents a BAC-derived segment. Bars to the left and right of centre line represent DNA copy number losses and gains, respectively. The green and red lines to the left and right side of the central line represent signal intensity $\log_2$ ratios of -0.5 and +0.5 respectively. Paired normal (N12) and HGD (Oral12) samples are shown in the blue box.

Figure 2. Genetic changes observed in each histological group on the 3p arm. Significant differences in the frequency of 3p alterations between LGDs and HGDs ($p = 4.60 \times 10^{-4}$) and between HGDs and OSCCs ($p = 0.025$) were observed.

Supplementary Figure S1. Copy number profiles in four normal human individual. *A*, two regions with overlapping BAC clones revealed copy number variations in the normal human

population. Genes listed on the left were identified as candidate TSGs from previous studies. These regions were excluded from analyses. *B*, gene tracks from UCSC May 2004 assembly displaying two regions that were excluded from our analyses. These regions harbor many previously discovered candidate TSGs.

Supplementary Figure S2. The detected altered region in each sample is shaded in pale green on the corresponding data points. Each data point represents one BAC-derived segment on the array. The raw data is displayed by *SeeGH*, whereas the detected change is displayed on the right side of the data by either *LSPHMM* or *aCGH-smooth* segmentation algorithm. Each MAR is marked with blue lines to show the upper and lower boundary of each MAR. The green and red lines to the left and right side of the central line represent signal intensity $\log_2$ ratios of -0.5 and +0.5 respectively. Data points to the left and right of centre line represent DNA copy number losses and gains, respectively. For *LSPHMM*, the left and right lines represent 100% probability for the data to be a loss or a gain, respectively.

Supplementary Figure S3. Integration of copy number and allelic imbalances in 3p14.1-p14.2. Allelic imbalances using three microsatellite markers, spanning region from 3p14.1 to 3p14.2 (harboring the *FHIT* gene) were previously assessed in all LGDs, HGDs, and OSCCs. A total of 88 samples yielded informative results at these microsatellite markers, and thus were integrated with its copy number status at the assessed loci. 32.9% of the cases showed concordant LOH and copy number loss, while 30.7% of the cases exhibited copy number neutral-LOH, and 12.5% of the cases demonstrated copy number loss without detection by LOH. 23.9% of the cases showed concordant copy number neutral and retention at LOH.
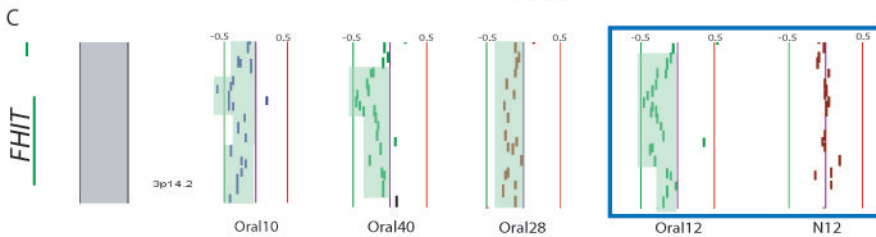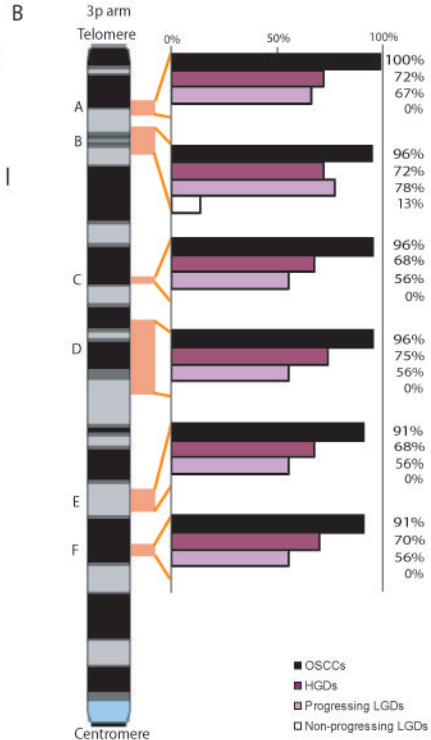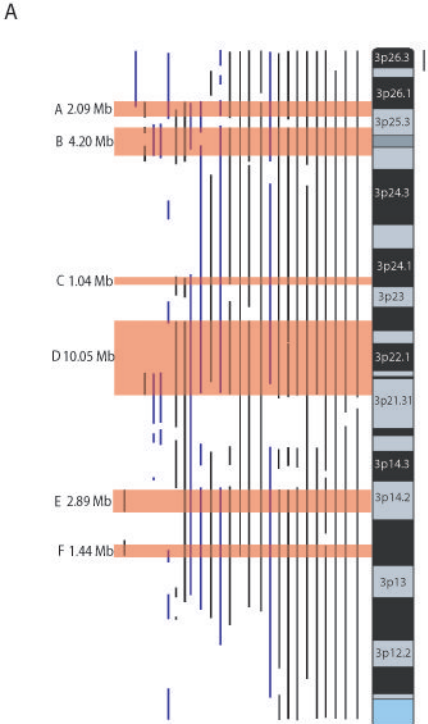
Table 1. Summary of recurrent minimal altered regions identified in 47 high-grade dysplasias using 3p tiling-path array CGH.

Region	Chromosome Band	Proximal flanking clone*	Start bp coordinate	Distal flanking clone*	End bp coordinate	Size (Mbp)	Genes[†]	Genes associated with with cancer[‡]
A	3p25.3-p26.1	776H3	7152555	629B9	9237884	2.09	7	*GRM7, RAD18, SRGAP3*
B	3p25.1-p25.3	525N21	10413544	464P4	14696702	4.28	28	*ATG7, TIMP4, PPARG, WNT7A, XPC*, *RAF1*
C	3p24.1	775G14	30598681	48 E16	31812073	1.21	4	*TGFBR2*
D	3p21.31-p22.3	606K24	36367289	494P19	46609837	10.24	95	*MLH1, LRRFIP2*, *ITGA9*, *CTDSPL*, *PLCD1, DLEC1*, *MyD88*, *VILL*, *ACVR2B*, *GORASP1*, *AXUD1*, *CX3CR1*, *CCR8*, *CTNNB1*, *CCK*, *CCBP2*, *TMEM158*
E	3p14.2	638K20	59594324	126N04	62484320	2.89	5	*FHIT*
F	3p14.1	413B3	67349894	259O22	68785539	1.44	2	-

[*]All the listed human BAC clones were selected from the RPCI-11 library.

[†]RefSeq Genes (release 25) according to UCSC May 2004 assembly.

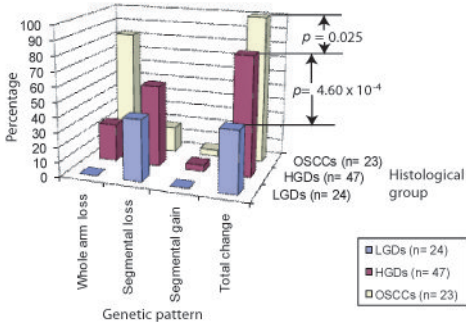[‡]Genes bolded are candidate TSGs in various cancer types.

Table 2. Pattern of 3p alterations in oral dysplasias and OSCCs.

	Whole arm loss*	Segmental alterations*	No genetic change*	Mean size of loss (Mbp)
LGDs† (n=24)	0	37.5% (9)	62.5% (15)	41.9
Non-progressing LGDs† *(n=15)*	*0*	13.3% (2)	86.7% (13)	0.78
Progressing LGDs† *(n=9)*	*0*	77.8% (7)	22.2% (2)	53.7
HGDs (n=47)	25.5% (12)	55.3% (26)	19.1% (9)	67.5
OSCCs (n=23)	82.6% (19)	17.4 % (4)	0	83.5

[*]Number of cases are given in parentheses.

[†]LGDs are categorized into non-progressing and progressing LGDs (see Materials and Methods).