

Translation of an STR-based biomarker into a clinically compatible SNP-based platform for loss of heterozygosity

Heather D. Kissel^{a,b,*}, Patricia C. Galipeau^{a,b}, Xiaohong Li^{a,b} and Brian J. Reid^{a,b,c,d}

^aHuman Biology Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

^bPublic Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

^cDepartment of Medicine, University of Washington, Seattle, WA, USA

^dDepartment of Genome Sciences, University of Washington, Seattle, WA, USA

Abstract. Loss of heterozygosity (LOH) has been shown to be a promising biomarker of cancer risk in patients with premalignant conditions. In this study we describe analytical validation in clinical biopsy samples of a SNP-based pyrosequencing panel targeting regions of LOH on chromosomes 17p and 9p including TP53 and CDKN2A tumor suppressor genes. Assays were tested for analytic specificity, sensitivity, efficiency, and reproducibility. Accuracy was evaluated by comparing SNP-based LOH results to those obtained by previously well-studied short tandem repeat polymorphisms (STRs) in DNA derived from different tissue sources including fresh-frozen endoscopic biopsies, samples from surgical resections, and formalin-fixed paraffin-embedded sections. A 17p/9p LOH panel comprised of 43 SNPs was designed to amplify with universal assay conditions in a two-step PCR and sequence-by-synthesis reaction that can be completed in two hours and 10 minutes. The methods presented can be a model for developing a SNP-based LOH approach targeted to any chromosomal region of interest for other premalignant conditions and this panel could be incorporated as part of a biomarker for cancer risk prediction, early detection, or as entry criteria for randomized trials.

Keywords: Pyrosequencing, CDKN2A, TP53, LOH, biomarkers, translational research, Barrett's esophagus, esophageal cancer, esophageal adenocarcinoma, cancer risk prediction

1. Introduction

Loss of heterozygosity (LOH) assessed with short tandem repeat (STR) polymorphisms has been shown to stratify patients with the premalignant condition Barrett's esophagus (BE) into low and high risk for progressing to esophageal adenocarcinoma (EA) [16]. BE is a premalignant condition in which the normal lining of the esophagus is replaced by a metaplastic columnar lining [55]. Most cases of BE are benign and this neo-

plastic, mucus-secreting tissue may in fact be a beneficial adaptation to chronic acid reflux in many persons [10,20,34,35,54]. Patients with BE have a 12- to 30-fold increased relative risk of developing esophageal adenocarcinoma (EA) [18], yet the overall probability that a patient with Barrett's esophagus will progress to cancer is around 0.7% per year [53]. EA is increasing in incidence and has become an epidemic in Western countries; however, most patients with BE die of unrelated causes [2,30]. The current standard of care for patients with BE is endoscopic surveillance, and surgical intervention by esophagectomy is frequently recommended for high-grade dysplasia. However, this procedure has an 8–23% associated mortality in high and low volume institutions, respectively, based on data from the Medicare database [6]. Therefore, imple-

*Corresponding author: Heather Kissel, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, C1-157, Seattle, WA 98109, USA. Tel.: +1 206 667 5776; Fax: +1 206 667 6132; E-mail: hkissel@fhcrc.org.

mentation of reproducible biomarker assays to stratify patients at high risk for future development of cancer that can be implemented in a clinical setting is needed.

Progress has been made in developing a panel of mechanistic-based somatic genetic markers that identify patients with BE who are at the highest risk of progression to adenocarcinoma [40–42]. In a 10-year prospective cohort study most consistent with an EDRN phase four study design [50], a biomarker panel, targeting measures of chromosome instability (LOH and ploidy abnormalities), improved prediction of EA in persons with BE [16]. Patients with at least three biomarker abnormalities, including 9p LOH, 17p LOH, and DNA content of tetraploidy or aneuploidy had a six year cumulative incidence of EA of 79.1%, as compared to 0% in patients with no biomarker panel abnormalities. This is one of the largest longitudinal studies of premalignant disease to evaluate somatic genetic abnormalities as predictors of progression to cancer, and its results are consistent with earlier independent studies of 17p LOH and aneuploidy in BE [11,52].

The 10 year prospective cohort study of patients with Barrett's esophagus began in 1995 using STR loci for LOH detection [16]. Following the roadmap put forth by The National Cancer Institute Translational Research Working Group (<http://www.cancer.gov/aboutnci/trwg>), steps to move from discovery research into a clinical application to reduce cancer incidence, morbidity, and mortality require translating a biomarker from a research assay into a platform adapted for clinical compatibility. In the present study, we sought a technology that could assess measures of chromosomal instability and 1) be compatible with high-throughput technologies, 2) amplify fragmented DNA using standard assay conditions for all polymorphisms, 3) have streamlined assays with minimal sample handling to reduce sample contamination, and 4) be robust using low amounts of DNA. Current state-of-the-art technologies using single nucleotide polymorphisms (SNPs) are amenable to more robust and high-throughput methods, and SNPs are present in higher densities than STRs throughout the genome, including intragenic loci. In addition, SNP assays can be designed to amplify small regions surrounding a SNP, making these polymorphisms more robust for use with formalin-fixed, paraffin-embedded (FFPE) or degraded DNA. We selected pyrosequencing because it can assess LOH, mutation, and methylation, it has been shown to be quantitative and reproducible, and already has regulatory approval for some indications [1,23,36]. In this study we developed a technically-robust panel

of SNP loci to replace the STR loci used for LOH on chromosomes 9p and 17p in our 10-year longitudinal study in order to overcome some of the challenges encountered in translating STR biomarker assays to the clinic. We then compared LOH as assessed by the SNP panel to LOH detection by STRs in all sample types.

2. Results

2.1. SNP selection

Chromosomal regions were selected to match the location of 12 STRs on chromosomes 9p and 17p as reported in a prospective cohort study of CDKN2A/ARF (9p21) and TP53 (17p13.1) tumor suppressor genes [16]. Initial tests of pyrosequencing were created for nine SNPs in collaboration with Biotage (Uppsala, Sweden). Subsequently, defined criteria were used to select and test an additional 34 SNPs as shown in Fig. 1 and these 34 SNPs were subject to the extensive additional testing to ensure all 43 SNPs were of the highest quality. Details of the final 43 SNP assays are provided in Supplemental Table S1.

2.2. Specificity of genotype

Genotype specificity was defined as the ability to accurately detect only a single, unique polymorphism. We determined specificity in three ways: 1) Each SNP assay sequence was required to align to a single genomic region using an in-silico tool (BLAT: BLAST-Like Alignment Tool) [21] during the primer design. 2) A single, pure amplicon was required to be detected in wet-lab tests of 5ul PCR product run on 2.5% agarose gels. 3) During genotyping, a set of non-polymorphic nucleotides within the sequenced region near each SNP was used as internal controls by the Biotage software to calibrate the sequence flanking the SNP and to determine whether the genotype will pass or fail (Fig. 2A). Using the two nucleotides sequenced at the end of each run, we determined the relative luminescence unit (RLU) ratio of these control nucleotides in each sample in order to determine if the assays designed amplify the sequence specifically. In theory, this ratio should be constant regardless of the SNP genotype being assayed or the input DNA used. The RLU concordance in the control nucleotides in 32,768 passing sequences across all assays and input DNAs in this study had an average R^2 value of 0.988 (range 0.9736 to 0.9975, in all passing assays, see example Fig. 2B).

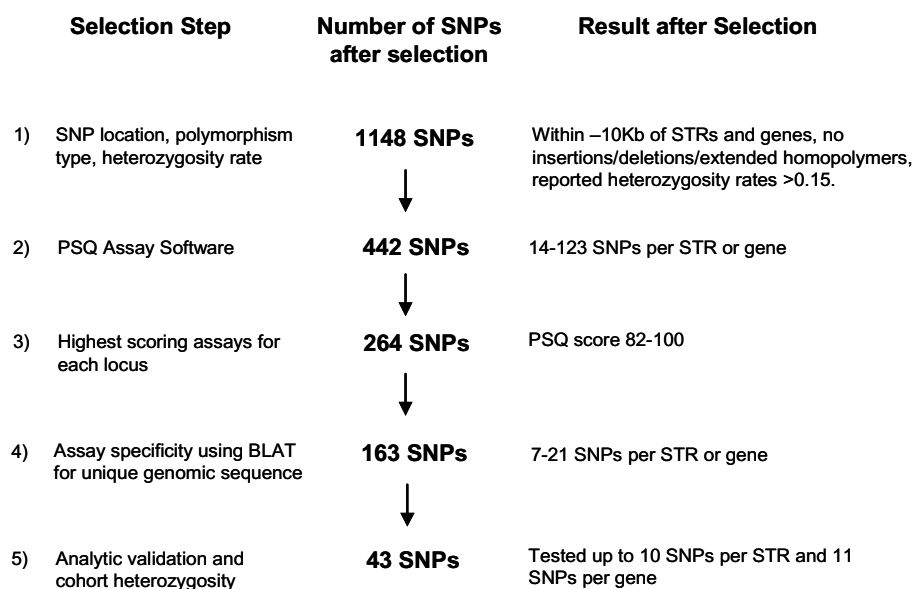


Fig. 1. Steps in SNP selection for SNP-based CDKN2A/TP53 LOH panel. SNP selection process for pyrosequencing-based LOH panel. Step 1) SNPs were selected from www.ensembl.org (release versions 36 to 38) to be ± 10 Kb from the current chromosomal location of the 12 STRs, TP53, and CDKN2A as reported in the NCBI SNP database, <http://www.ncbi.nlm.nih.gov/projects/SNP/>. SNPs with extended homopolymeric regions were excluded. Homopolymers are defined as having multiple nucleotides flanking the SNP that are the same nucleotide as the one of the bases in the SNP itself. SNPs with reported heterozygosity rates below 0.15 in Caucasian populations were excluded, and SNPs with unknown heterozygosity rates were further evaluated. Step 2. The flanking DNA sequence for each polymorphism was entered into the PSQ Assay Design Software (version 1.0.6) from Biotage (Uppsala, Sweden) to determine SNPs that make optimal pyrosequencing assays using a stringent Allele Quantification setting. The software ranks each assay on a scale of 0 to 100, with 100 being most optimal. Step 3. To limit the number of assays required for analytic testing, only the highest scoring SNP assays were used (range 82–100). Step 4. To determine assay specificity, the resultant DNA sequences surrounding each SNP generated by the PSQ software were entered into BLAT (Blast-like analysis tool) software to BLAST (Basic Local Alignment Search Tool) against the genome (<http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>) starting with the sequence of the forward primer and ending with the reverse primer. Step 5. 163 SNPs were available for analytic testing and 70 were tested to obtain 34 SNPs. These were combined with the nine SNPs created by Biotage, resulting in the final panel of 43 SNPs.

2.3. Efficiency of assay

Once each SNP assay was determined to be specific to a unique polymorphism, we evaluated the assay efficiency as measured by the genotype pass rate in the following ways: 1) PCR reactions at different annealing temperatures, 2) Different input DNA sources, and 3) Comparison of SNP assay efficiency to the gold standard STR efficiency on sub-optimal FFPE DNA.

The efficiency of this protocol at the optimal PCR annealing temperature of 57°C was compared to PCR reactions at annealing temperatures of 54°C and 60°C. All SNPs tested for the panel were selected to amplify optimally using the same reagent concentrations and PCR reaction conditions such that the entire SNP panel could be run simultaneously using a 96-well plate in a single thermocycler at 57°C annealing temperature. Eight constitutive control gastric DNAs were amplified using 34 SNP assays for a total of 272 genotypes at each PCR annealing temperature. 248 out of 272 (91%) genotypes passed at the optimal temperature of 57°C.

Table 1
Comparison of sample type and genotype pass rate

Sample type	# Samples	Genotypes	Pass rate
Constitutive control	155	16118	94%
Endoscopic biopsies	279	12556	93%
FFPE	46	2076	67%
Surgical resections	138	943	79%

In contrast, the pass rates at sub-optimal temperatures of 54°C and 60°C were 192/272 (71%) and 179/272 (66%), respectively.

For all 43 SNPs, we assessed assay efficiency using 618 DNA samples from four sources with different input DNA concentrations including 1) constitutive control DNA from fresh-frozen gastric biopsies, 2) DNA from fresh-frozen endoscopic biopsies from BE, 3) DNA from FFPE BE biopsies, and (4) DNA from surgical resections. Overall genotype efficiency varied between sample processing methods (Table 1). Surgical resection samples were genotyped during template preparation optimization and pass rates of these sam-

Supplemental Table S1
Detailed information about the SNP assays

CHR	STR	Assay ID	SNP ID	Basepair Position	Distance to nearest STR ¹	Dispensation Order	PSQ Score	Reported Het Rate ²
9	D9S935 and D9GATA62F03	ASY114	rs18622779	5,190,127	inside STR	GACATCGAT	100	0.458
9	D9S935 and D9GATA62F03	ASY136	rs10491650	5,193,054	inside STR	ATCTGATCA	95	0.467
9	D9S925	ASY270	rs10810945	18,278,212	inside STR	CGATCTGCA	95	0.483
9	D9S925	ASY170 ^d	rs10963557	18,279,897	inside STR	GACATGTAT	92	unk
9	D9S925	ASY131	rs7862129	18,280,410	inside STR	GCTATCAGT	97	unk
9	D9S925	ASY117	rs10738506	18,280,419	inside STR	GATCAGTGA	100	0.300
9	D9S925	ASY247 ^d	rs13285950	18,281,228	inside STR	CTGACGACT	83	unk
9	CDKN2a	ASY484 ^d	rs1134871	21,852,897	104,854 bp upstream	GTAGTCACT	64	unk
9	CDKN2a	ASY233 ^d	rs2518719	21,960,427	intragenic	GTACATGAC	91	0.233
9	CDKN2a	ASY486 ^d	rs2811708	21,963,422	intragenic	CGACTGAC	93	0.666
9	CDKN2a	ASY235 ^d	rs4074785	21,971,583	intragenic	CGACTGCTA	91	unk
9	CDKN2a	ASY238 ^d	rs3731217	21,974,661	intragenic	CGTCACTGA	90	0.233
9	CDKN2a	ASY236 ^d	rs3731211	21,976,847	intragenic	GATGATCGA	91	0.367
9	CDKN2a	ASY245 ^d	rs10757262	21,977,874	intragenic	GATGACAGT	90	0.200
9	CDKN2a	ASY479 ^d	rs2811711	21,983,964	intragenic	ATCAGTACT	100	unk
9	CDKN2a	ASY232	rs3218012	21,988,660	4,170 bp downstream	GATCTAGAT	96	0.417
9	CDKN2a	ASY493 ^d	rs1063192	21,993,367	8,877 downstream	GCTCGATAT	95	0.534
9	D9S1121	ASY264 ^d	rs7867884	25,391,353	inside STR	GACTGCATG	88	unk
9	D9S1121	ASY142	rs10757533	25,393,652	inside STR	CGTGTCTAG	98	unk
9	D9S1121	ASY120 ^d	rs10812234	25,393,950	inside STR	ATCAGTCTG	100	unk
9	D9S1118	ASY122	rs4879542	31,911,530	inside STR	GCATAGCAC	100	0.500
9	D9S1118	ASY138 ^d	rs10970600	31,911,724	inside STR	TTAGCTGTA	95	unk
9	D9S1118	ASY125	rs16917529	31,914,600	inside STR	AGTCAGTAG	100	0.208
17	D17S1298	ASY100	rs8080482	3,611,865	1,434 upstream	GCATCTGCT	100	unk
17	D17S1298	ASY158	rs2731739	3,613,611	68 downstream	TCGTCAGTG	98	0.233
17	D17S1298	ASY101 ^d	rs966004	3,616,376	3,077 downstream	GATGCTCTC	100	0.273
17	D17S1298	ASY203 ^d	rs8502743	3,617,632	4,089 downstream	CTGACTGTG	91	0.292
17	D17S1537	ASY103	rs2106233	5,841,753	1,694 upstream	GATCGACA	100	0.208
17	D17S1537	ASY282 ^d	rs11654718	5,842,766	681 upstream	ACTGTGATG	91	0.500
17	D17S1537	ASY167	rs9907608	5,844,015	377 downstream	CTGTCTGAG	96	unk
17	TP53VNTR and TP53DI	ASY480 ^d	rs2270341	7,421,271	91,192 upstream	GTAGTACTG	96	0.433
17	TP53VNTR and TP53DI	ASY481 ^c	rs858526	7,440,107	72,356 upstream	CTGGGAGTA	83	0.417
17	TP53VNTR and TP53DI	ASY477 ^c	rs858521	7,470,872	41,591 upstream	TGCGATAGC	98	0.417
17	TP53VNTR and TP53DI	ASY220 ^d	rs12951053	7,518,132	intragenic	AGTGCATAC	100	0.167
17	TP53VNTR and TP53DI	ASY222	rs2909430	7,519,370	intragenic	AGTCAGTAG	90	0.226
17	TP53VNTR and TP53DI	ASY482 ^d	rs2287499	7,532,893	1,252 downstream	ACCGTAGCT	100	0.183
17	TP53VNTR and TP53DI	ASY478 ^c	rs2270517	7,691,572	159,931 downstream	ATCAGCTCT	100	0.345
17	D17S786	ASY107	rs726680	8,752,394	111 upstream	CAGTGATAC	100	0.316
17	D17S974	ASY110	rs12601954	10,458,827	609 upstream	AGTCAGTGA	91	unk
17	D17S974	ASY191 ^d	rs8079868	10,461,425	1,786 downstream	TGCTAGGCA	100	0.273
17	D17S1303	ASY209 ^d	rs2692343	10,801,344	1,001 downstream	GTAGTGTGA	91	0.417
17	D17S1303	ASY210 ^d	rs9907650	10,801,868	1,525 downstream	GTAGCAGAG	90	unk
17	D17S1303	ASY281	rs2247382	10,801,962	1,619 downstream	GACATCTGA	82	0.400

Experimental Het Rate ³	Amplicon Sequence (Sequence Analyzed)	Forward	Reverse	Sequencing
0.355	ACAATCJAATGAAGTATTTCAGCATACA	B ⁴ .CCCAATGCCTTTGGTATGCT	GGAGACAACTGGGGATCTTTCAA	TTTGACTCTTACACGAGG
0.352	TCGGAATTCAGGACTCAGAAATTTGTTCT	B-TCAATGGGATCTCCATAGAACA	TGTGCAGAAAATCAACTCTTCTC	CAATTTCTGTCACTTGCTA
0.344	GAATTCGCAATACAGTGAATTAAGACAGTT	GAAGCTGCTGGCCAGTATAAA	AGACATGACCTGGGATTTTCTC	CAGTATAAAGCAGACCAACGA
0.140	ATGTTGATTTGTTTTCATGTTGTTTTC	B-TCAAGACAGGCTGGAACCTTCA	CACACTGCTGGTACCTTGCAAT	GGTTACCTTTGGTCAATC
0.204	CTAATCGGGTGAACCTCAGATTGGCTCAGACA	B-CAAAGCAGTGAAGGGGACAGT	CATCTTTGGTGTCCATGTGAT	GTGGTGTCCATGTGAT
0.457	ATTCGTTGATCTAGTGGAACTCAGATTCG	B-TTGCTGAGCCAAATGTGAGTTC	ATTAGCTTTGGGGCCAGGA	CACATCTTTGGTGTCC
0.161	TGTAAGAACTAAATGAAGCTTGATACCC	GGGCTCAGAAATTAATGAGACTTA	B-GGGTATCAAGCTTCATTTAAGTTC	AAATAAGTAACTAAATGCC
0.290	ATATTCGACTGAGAAAAA	B-TTGAATCTTTCTGTGCTTCC	ACCAGAAACAGGTTTTCAGAAA	ACCACTTACTAGTTGATTA
0.228	TTTCTGCTCACAACTCAACTCTGACCC	B-TGGAAACTACCCGCATAAAGC	CTAATTTCTCATTTCATACCA	TTTGCATCTGGTCTC
0.344	AGAAATCAGGGAAGAAGA	B-GGATTTCAAGCCATCATCTTCC	GGGAGTTGGCTTTGATGCTCA	TTTTGACAAATTTAATGG
0.207	GGATTCCTCAATTTCCCTTACATTTGCTTT	B-TCCATGCATCCCTTTAATTTAGA	TTTCAGGTAACAAGTAGGGTG	TTGGTGTGAACATGACACG
0.256	GTGAGTGAAGCCGAATACATATTATT	AACGAGTGTAAACCTATCAGTGA	B-TACGCATGTGCTTTAGAGTCT	AACCTCCACTCCGAC
0.355	ATGATTTGGACATTTAACAATAGATGTTCA	B-TCTCAGAAATTTAAAGAAACAGC	GCATGGAGTAAATTCAAAGTGTC	AAAATTAAGTGTCCG
0.247	AAATTTCAAGTGGCCCAATGGGACACAAT	B-TCCCAATGGGCCACTGTGA	TGCTGGGGTCCCAAGAACT	AAACACTTGAAGGAGATACT
0.255	TCGGTACTGCG	AAGGGAGGTCCAGGTTCA	B-TGGGAGGCTTCCCAATTC	AAGGGAGGTCCAGGTTCAA
0.465	AAATTCAGAAAATTTATCCCTGTTAAAA	CCCTGCCTGAATGAATTT	B-GGCCAAAAACCTGAGTTAATGA	GAATGAATTTGCTGATCTA
0.452	TCGATTTGATTTAAGGG	GCTAGTTGATTAATCTGGGTCAT	B-TGGGATCCACAATAAATTTCTT	GAATTTGCTTAATGACAA
0.489	ATTCCTCCATGAATAACCCCTTGGTTA	B-TCAAGTTAAACAAAGGGGTTTAT	CTGTTGAGGCTGTTAACTCTC	TTTTTATAAATCATGGAGT
0.493	CTGTTGATGACTTTTAAACCACTGAGTAA	B-TAGCTTTTTAGTTGCCAAGAATA	TGGATCAGACAACCTTTTTTCTA	AAATCTGTCACTTGTCTC
0.478	TCGCTGCTGCTGACCTGCCCAAGCGAAA	TGTGAAAGAAAGTGGGACTAGTG	B-CCTGGCCAAAGATGCTTTATACT	CACCCATCTCTTATATC
0.544	CTATGCAACATAGAGAAAGCAGAAATGAAG	TGTAGCCCTCTGTTTTCTACTTT	AACCAAGCAATGACAATCATCT	GGCAATGACAATCATCT
0.183	GTATTTGTAAGCTGAGAAATGGCCAGA	B-GCTAGTGTCTCAGACCTTGTGAT	CATGCGTGTCTTGCATATAG	AAAGTAGAACCAAGAGACA
0.322	GTTTGGTGGTGGAGGCTTTTGGACCT	B-GGGTATAGAGAGGGTCAAAAAGT	AGCTGCCATCTTGTGAATAAT	GTGGTCAACATGGAAG
0.402	CAATTCGCTGTTCCGAGGCTGAGTCTAGCG	B-ACTCAGCCTCGGGAACAGC	CTAGCCCTCGTGTGACCC	AGCAGCCTCTGAGTCTC
0.226	CGTTTCGTTTGGAGATGAAAACAGAAATGA	TGGCAACCAAAAGACCTACATTC	B-CGGCAGTCTCTGAGTCAATCT	GCCTCTCTTATAAATGG
0.304	ATCTTTCTCCGACTCTGATCTCTGCTG	B-GGAGGCAAGGAGATCAGAGTC	GGCTGCAGCATAGCAAG	GCATGCAAGCTTCAACA
0.297	GTGGCTTTGTTGAGATGAAAACAGAAATGA	GTCCCTGACTCTCAAATTTATTC	B-CAGGAGGCTATTACCATCAAGAG	CCTTCCATCATCTGCTC
0.283	TAATTCAGAAAGCATAGTTACTCATTTCTG	AGGGGAGCAAGCAGTACAGGT	B-CCCTTTCTTCCCAAGAAAGTGA	GGTCCAAATCTGCACT
0.386	CTTCGATTTGGAAGTGGATCTTGTAGCCC	AGGCTCACTGAGGGCTACAA	AGCTGCCTCTGTGAGTGAAGT	CCTCTGTGAGTGAAGTCCA
0.301	TGTTTCGAGAAAGGTGATGAGGACGCTGGTG	AAAGGTGAGTTTGAACCCCAAAAT	B-AAAGTCACTCAGGCAAGAAAGCA	CCAAAATGTTGAAAGAAA
0.576	ATATTAAGTGGATCTTGAAGCT	B-CCCTTCTAGCCACACCTAA	GAGGGCAGTCTTGTGAAATAG	GTGGCAGGGGAACCA
0.564	TGCGTAAGTAAATAAGGAA	B-CGGACTCAGAGTGGTCTTAGC	CAGGCACACGACCCCTTG	AGGCCACGACCCCTT
0.447	TCGTAAGTAAATAAGGAA	CCAGCCGGTGTCAATTTT	B-CCTTGACCTGTTGGAGAGAA	CCCTCTAGCTGAAA
0.141	GTTCCTACTACTACCCTCCACCTCTC	B-GAGTGGGAGCAGTAAAGAGATTC	ACTTCCACCCTGCACAC	CACCTCTTACCAGTTTCT
0.269	GTTTCGTAAGCACTCTCCGCAACCCACTAG	B-GACGCCACTCTCTAGCTCG	GAAAGTCAGGGCAAGTGAACAG	CAGCAAGAAACCAACT
0.135	CCGCTAGCTCTGGGACACAG	CAACTCTGTTCCAGGGGAGTG	B-ATCTCTGGCTGGCTAGCT	AACTGGTCCCTCTC
0.426	TCGCTTCACTACCTTGA	ATGGTGGAGCTTGTCTTGAAG	B-GCCGGGCTGCAAGTGTAGA	CAATGGCTCCTACACC
0.355	ATGTAATAAACCAGAAAGTCTGGAT	ACCACCTTGAGTGCAATAACA	B-ACCTCCTCTTTGTCAACAG	CCTCTGGAGTGAATAAC
0.247	GGTTTCGTTGGAGCTCTTTTCTCACCCT	CGCTTCTGGTGCATTTTC	B-CCTAAAGCCCAAGTGTGA	CTGGTGTAAAGGACT
0.247	GGCTTCGTTGGAGCTCTTTTCTCACCCT	AGGGCTTCAACAGGAGA	B-GTCACTGGCTGGAACCAACTC	TCAGAGGGAATAAAGG
0.441	TAGTTCTAGTTACGTTGATGATAAGTCTC	B-GGGTTTGTATAGGATTTGGA	B-GCTTAATGGAAACACTGAAT	TGGAACACTGTAATGAGTG
0.118	TAGTACGAGAGAGAAATCAAGAAAGATAATGA	B-CGTAGATCTTCCCTGATCTCAT	AACAATAGTCTTTGGGGCTACAC	AAGAAATGGGGCTATCTC
0.441	ACAATTCGAAAACAATCTGTTGGGGCTACAC	B-ATATCATTATGTAGCCCCAAC	AAGAAAGTCCATACAATTTGGTA	ATAAATGTGCTTTGAAAT

³ Reported heterozygous genotype frequency from www.ensembl.org, version 41

² Experimental heterozygous genotype frequency derived from cohort

¹ SNP assay designed in-house by Biotage

⁴ Homopolymer assay

⁵ Biotinylated primer

¹ Basepair distance is calculated from the beginning (upstream) or end of the STR (downstream)

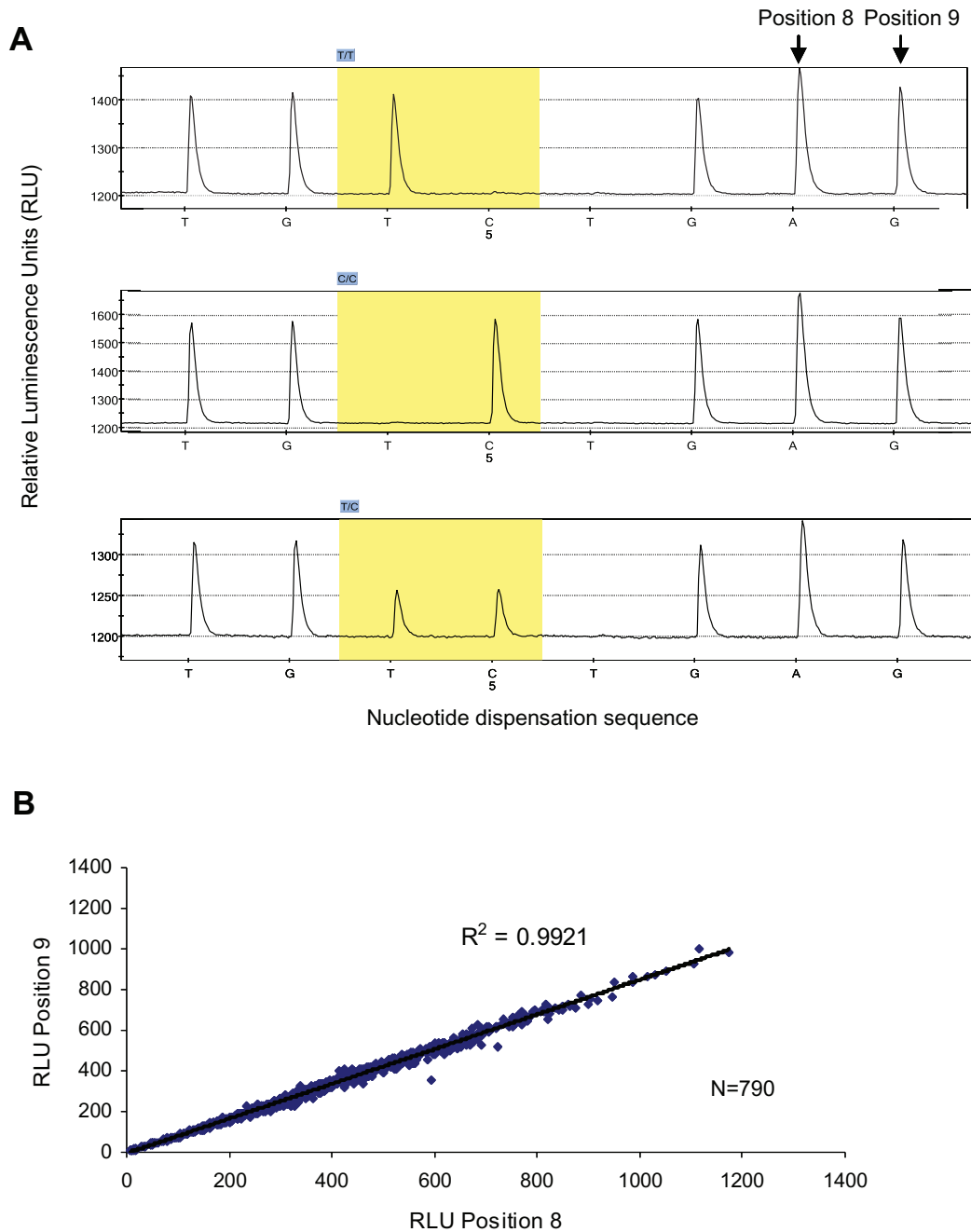


Fig. 2. (A) Pyrograms showing all three genotypes of the SNP (yellow). The RLU ratio at positions 8 and 9 does not change for different genotypes. (B) Scatter plot of peak heights at Positions 8 and 9. (A) Pyrograms from constitutive control DNAs showing peak intensity of each nucleotide sequenced for ASY167 for homozygote T/T, homozygote C/C, and heterozygote T/C. Positions 8 and 9 of the sequence are used as non-polymorphic internal sequence controls for each sample and can be used as a measure of specificity across all samples. (B) Concordance of the peak intensity ratio of control positions 8 and 9 for 790 genotypes of ASY167. (Colours are visible in the online version of the article at www.iospress.nl.)

ples reflect some variation in pyrosequencing reaction conditions.

Given the lower genotyping efficiency of the DNA

from fixed tissue, we sought to directly compare the genotyping success rate of STRs relative to SNPs when amplifying DNA extracted from fixed tissue. We am-

Table 2
Pass rate of STRs and SNPs in FFPE samples

STR locus	Passing STR results/Total samples	SNPs within STR region	Passing SNP results/Total samples
D9S935	34/46(74%)	ASY114	20/46 (43%)
		ASY136	38/46 (83%)
D9GATA62F03	34/46 (74%)	ASY114	20/46 (43%)
		ASY136	38/46 (83%)
D9S925	26/40 (65%)	ASY117	27/46 (59%)
		ASY131	30/46 (65%)
		ASY170	40/46 (87%)
		ASY247	32/46 (70%)
		ASY270	26/46 (57%)
D9S1121	15/27 (56%)	ASY120	35/46 (76%)
		ASY142	31/46 (67%)
		ASY264	31/46 (67%)
D9S1118	15/23 (65%)	ASY122	37/46 (80%)
		ASY125	38/46 (83%)
		ASY138	36/46 (78%)
CDKN2A (No matching STRs)		ASY232	27/46 (59%)
		ASY233	28/46 (61%)
		ASY235	40/46 (87%)
		ASY236	36/46 (78%)
		ASY238	33/46 (72%)
		ASY245	35/46 (76%)
		ASY479	43/46 (93%)
		ASY484	39/46 (85%)
		ASY486	37/46 (80%)
		ASY493	33/46 (72%)
Average pass rate	68%		73%

plified 46 FFPE endoscopic biopsy samples from 23 patients using five chromosome 9p STRs (D9S935, D9GATA62F03, D9S925, D9S1121, and D9S1118) and all 23 SNPs for chromosome 9. The number of STRs run per DNA varied from two to five, depending on quantity of DNA available per sample. For this set of FFPE samples, the average pass rate was 68% for STRs and 73% for SNPs (Table 2).

2.4. Reproducibility of genotype

We determined the genotype reproducibility among multiple PCR and sequencing runs. For any given SNP, pyrosequencing generates a signal for each allele. In a single DNA sample, the SNP allele frequency in that population of cells is quantified by the fraction of one allele over the sum of both alleles $\{A/(A + B)\}$. We tested reproducibility by comparing the variance of SNP allele frequency in each sample, across multiple PCR annealing temperatures and across different PCR reactions and pyrosequencing reactions run on multiple days.

For 34 SNP assays, we combined the results generated from the eight patients run in the temperature experiment described above and evaluated the coefficient of variation (CV%) of the allele fraction for each

passing genotype (AA, AB, BB). The mean CV% for each SNP among genotypes A/A was 1.6% (95% CI = 0–4.4%), A/B was 3.4% (95% CI = 0–8.1%), and B/B was 0.9% (95% CI = 0–2.5%). Thus, for any given genotype, the variance of allele fraction is very small, even with large variability in amplification conditions. Among the three annealing temperatures evaluated there was a total of 619 passing genotypes, and 614/619 (99%) of these genotypes matched across the different temperatures.

At the target annealing temperature of 57°C, we sought to test the genotype reproducibility among both independent PCR and pyrosequencing reactions on multiple days. Using a larger sample panel of constitutive control gastric DNAs from 93 patients with BE, each DNA was subject to two different locus-specific PCR reactions and both PCR products were sequenced on the pyrosequencing instrument on two different days, resulting in four genotypes per DNA. A minimum variation of allele frequency of no more than 5% from the expected genotype allele frequency is recommended for an assay to be considered efficient and reproducible (Biotage, personal communication). Homozygous genotypes should theoretically have an allele frequency between 0 to 5%, or 95 to 100%, for the first SNP allele sequenced, and heterozygous geno-

types should fall between 45 to 55%. Figure 3A shows a typical pattern of genotypes using the allele frequency of the first SNP allele sequenced in each assay. A subset of tested SNP loci (3 of 34) showed a shift from near 50% allele frequency for homozygous genotypes (Fig. 3B), or a shift away from 0% or 100% for heterozygous genotypes (Fig. 3C), but these genotypes were highly reproducible among all samples within that SNP assay. For each SNP, we calculated the average passing allele frequency for up to 372 genotypes (93 DNAs X 4 replicates). For one of the 34 SNP loci, only 70% of passing genotypes fell within 5% of the mean allele frequency. All the remaining 33 SNPs had tight genotype distributions: Per SNP, the percentage of passing genotypes that fell within 5% of the mean allele frequency ranged from 88–100%, with the average being 95% of samples falling within 5% of their mean allele frequency. This genotype reproducibility is consistent across the entire dynamic range of RLU intensity for both homozygous and heterozygous genotypes (Fig. 3D)

2.5. Sensitivity of assay

SNP assays that are efficient and reproducible must also be sensitive to low input DNA amounts, varying DNA quality, and mixtures of cells. LOH is a somatic genetic abnormality characterized by clonal heterogeneity. Thus a sensitive, quantitative genotype is required for accurate LOH assessment. To evaluate assay sensitivity, we performed two types of experiments: dilution experiments using both fresh-frozen and FFPE tissue, and experiments in which constitutive DNA is mixed with samples with known LOH from the same individual.

Samples were amplified using a subset of six SNPs; the three most robust and the three least robust assays, as defined by our temperature and genotype reproducibility experiments described above. Two fresh/frozen constitutive controls from two separate patients were used, with concentrations ranging from 1 ng to 0.002 ng (~143 cells to 0.3 cell equivalents, assuming 7pgDNA/cell). Each of the six SNPs was run in triplicate on the two constitutive control samples (6 SNPs x 3 replicates x 2 DNAs) and then repeated in a separate experiment. Replicate reactions were compared to determine the frequency of allelic drop-off and percentage of genotypes that gave passing pyrosequencing results. The percentage of passing fresh/frozen samples ranged from 98% for 1 ng to 61% for 0.002 ng (Table 3). In a separate experiment using eight FFPE samples, the

pass rates ranged from 96% for 1 ng input DNA to 77% for 0.06 ng (< 9 cells) using the same six SNPs. These genotypes were accurate down to 0.015 ng input DNA. Below this level, a single heterozygous sample had allelic drop-off. Homozygous samples showed no genotype mismatches among replicates across the entire range of input DNA concentrations.

Six patients previously identified to have aneuploid populations with known LOH on chromosome 9p or 17p were evaluated for informativity in three SNP assays. One patient was informative for SNP IDs 477 and 478, and two other patients were informative for one SNP each. In these three patients, aneuploid populations were flow-purified and DNA was extracted. Aneuploid DNA was quantitated and mixed with constitutive control DNA from the same individual prior to PCR in increments of increasing amounts of constitutive control DNA at 0%, 10%, 25%, 50%, 75%, 90%, and 100%. Mixing was based on absolute DNA amount and did not adjust for copy number changes in the aneuploid. The two SNPs were evaluated in each of the constitutive control and mixed samples in duplicate, using ~6 ng DNA per reaction. Allele frequencies were highly reproducible along the entire range of allele mixtures ($R^2 = 0.98$) for all 27 duplicates. Figure 4 shows example pyrograms generated from one of the SNPs in one patient.

2.6. Accuracy of SNP-based LOH

Accuracy is defined as the agreement of a genotype to the “true” genotype. Our panel is built to determine LOH, thus we compared the accuracy of SNP-based LOH using this panel, relative to the gold standard STR-based LOH. Using the distributions of allele fractions from constitutive samples for each assay as a reference, a probability model was used to define distinct genotypes (see methods). This model classifies LOH calls in biopsy samples into categories of LOH yes, LOH no, and LOH intermediate. It is desirable to have an LOH yes/no call for a clinical application; therefore, for LOH call comparisons between SNPs and STRs, we were conservative and binned all LOH calls into LOH yes and LOH no, with any LOH intermediate calls classified as LOH no. 1379/1513 (91%) informative STR LOH results matched the LOH result of an corresponding informative SNP. These by-SNP results were derived from 274 samples from 100 patients (see methods) that had at least one informative SNP and a physically associated corresponding informative STR, resulting in 1–25 STR:SNP comparisons per sample

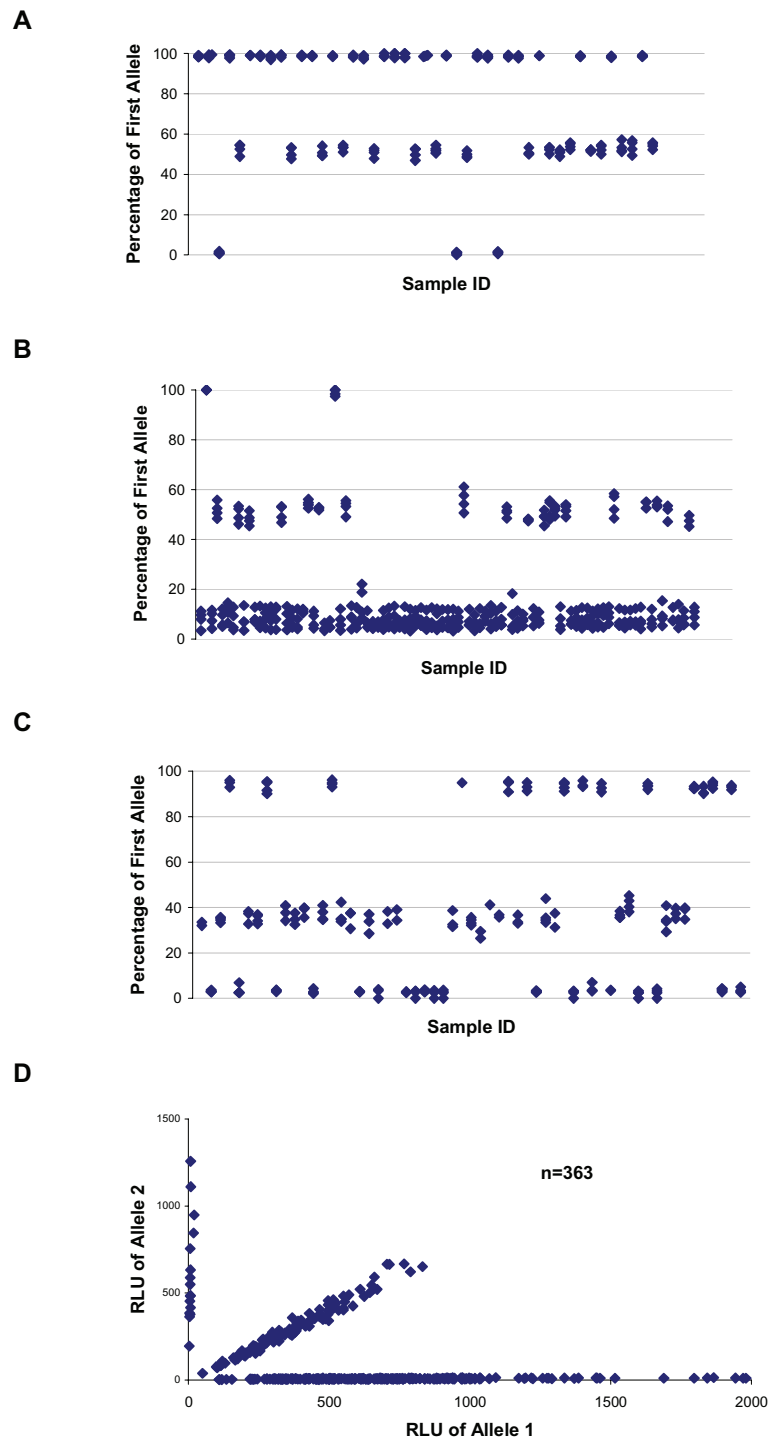


Fig. 3. Example of genotype reproducibility between independent PCR and pyrosequencing reaction genotypes. Allele frequency (y-axis) was determined by the pyrosequencing software. For each DNA sample, four data points are plotted one after the other, representing replicate PCR reactions which were both sequenced on two different days, shown in order of sample ID. (A) near 50% allelic ratio for heterozygotes, (B) shifted allelic ratio for homozygotes, (C) shifted allelic ratio for heterozygotes, and (D) distinct genotype clusters across homozygous and heterozygous samples. (Colours are visible in the online version of the article at www.iospress.nl.)

Table 3
SNP genotyping pass rate with decreasing input DNA from fresh-frozen and fixed tissue

DNA amount	1 ng	0.5 ng	0.25 ng	0.125 ng	0.06 ng	0.03 ng	0.015 ng	0.008 ng	0.004 ng	0.002 ng
Percent passing	98%	97%	96%	88%	81%	72%	61%	61%	61%	61%
Fresh/Frozen DNA	<i>n</i> = 72	<i>n</i> = 72	<i>n</i> = 72	<i>n</i> = 72	<i>n</i> = 72	<i>n</i> = 72	<i>n</i> = 36	<i>n</i> = 36	<i>n</i> = 36	<i>n</i> = 36
Percent passing FFPE DNA	96%	92%	90%	81%	77%	N/A	N/A	N/A	N/A	N/A

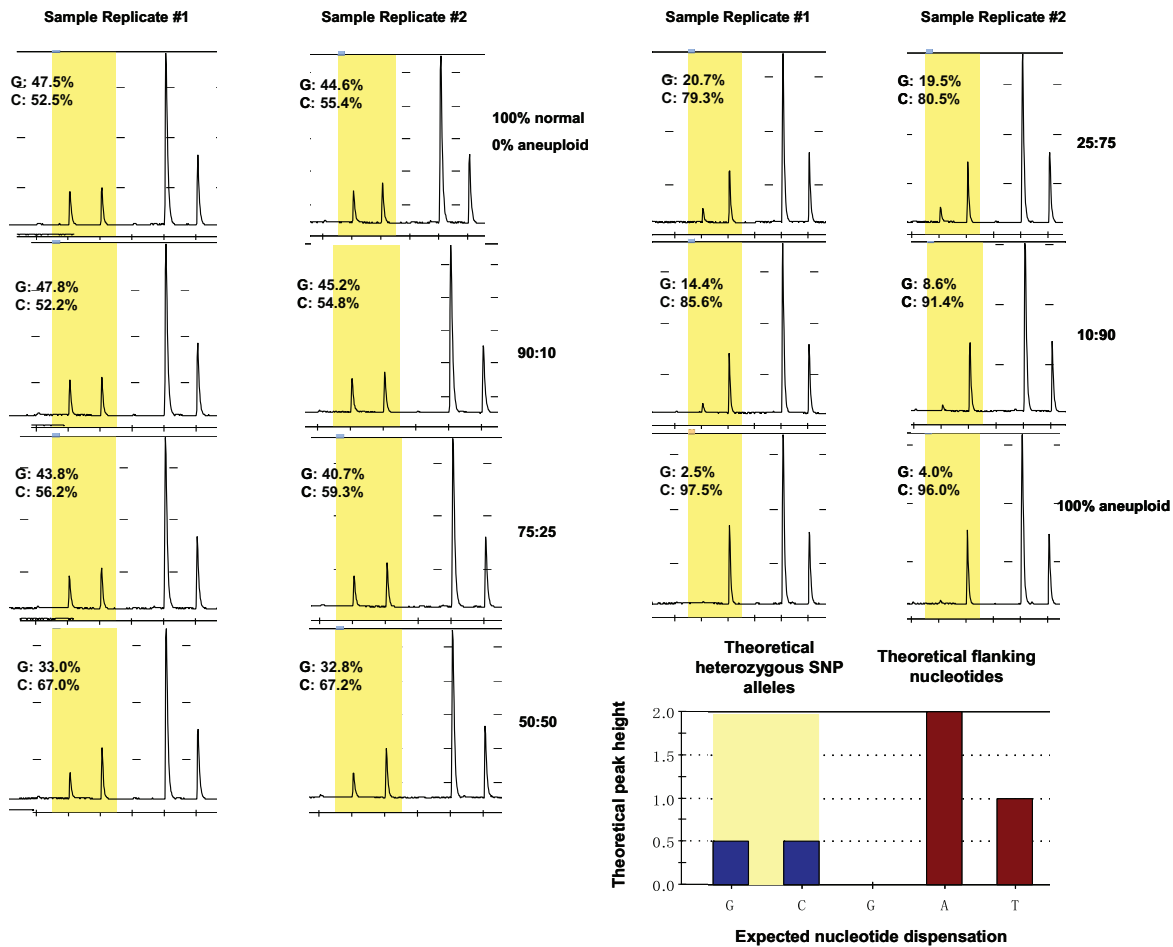


Fig. 4. Example of pyrograms from mixing experiment. Pyrosequencing pyrograms of constitutive control and flow-purified aneuploid DNA was quantitated and mixed at 0, 10, 25, 50, 75, 90, and 100% aneuploid. The first pyrogram shows that the results are very similar to the theoretical peak heights for a normal G/C heterozygote and the flanking sequence, as depicted in the final bar graph. Mixtures were PCR amplified for pyrosequencing at an informative SNP (replicate #1 and #2). LOH in the aneuploid resulted in loss of the “G” allele. (Colours are visible in the online version of the article at www.iospress.nl.)

(average = 5.5) and yielding a total of 1513 informative STR:SNP comparisons. For each informative STR locus (1–6 informative STRs per sample, average 2.5), 1–8 SNPs were informative (average 2.2), resulting in a total of 696 by-STR locus comparisons. 643/696 (92%) comparisons between an STR and at least one informative associated SNP gave the same LOH call in these 274 samples.

To achieve a by-chromosome LOH call, we used 225 fresh-frozen endoscopic biopsy samples from 84 patients (see methods). For chromosome 9p, 197/225 samples were evaluated for five STRs and 13 SNPs, excluding CDKN2A intragenic SNPs which do not have an associated STR. Regardless of physical location on the arm, STR and SNP LOH results agreed by chromosome arm in 175/197 (89%) of samples. For chro-

Table 4
STR:SNP comparison of by-chromosome LOH calls

Sample type	# Samples with LOH results for STRs & SNPs	# SNPs	Average informative SNPs/sample	# STRs	Average informative STRs/sample	Samples in agreement by whole chromosome arm
Endoscopic biopsies						
Chromosome 9p	197	13 ^a	3.5	5	2.7	89%
Chromosome 17p	221	20	4.9	6	3.1	87%

^a9 SNPs intragenic to CDKN2A were not included because there were no intragenic STRs evaluated.

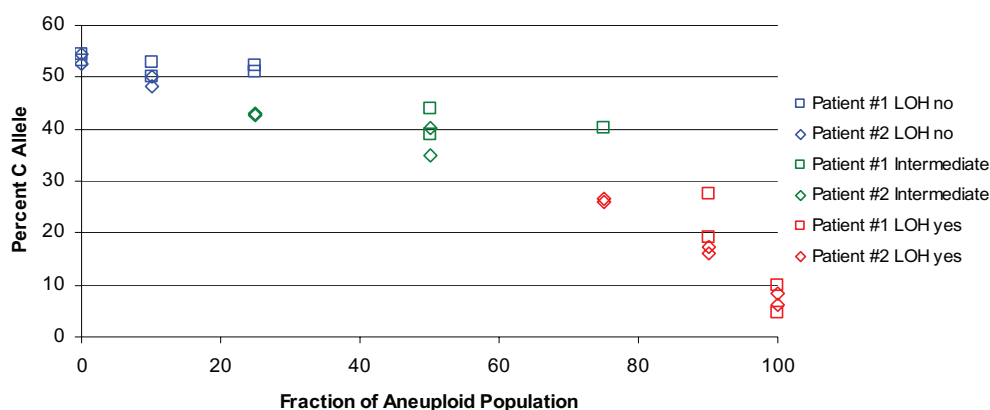


Fig. 5. Percent allele and LOH calls for constitutive control:aneuploid mixtures in two patients. Scatter plot of the mixing experiment using a constitutive control and paired aneuploid DNA for two different patients for one assay with LOH calls based on the statistical algorithm. The squares are patient #1 and the diamonds are patient #2. The blue is LOH no, the green is LOH Intermediate, and the red is LOH yes. (Colours are visible in the online version of the article at www.iospress.nl.)

mosome 17p, 221/225 samples were evaluated for six STRs and 20 SNPs. STR and SNP LOH results agreed by chromosome arm in 193/221 (87%) of samples (Table 4). By-patient, chromosome 17 (20 SNPs) had at least one informative SNP in 99% of patients (83/84) and chromosome 9 (13 SNPs) was informative for 88% of patients (74/84). When CDKN2A SNPs (10 SNPs) were included in the informativity calculation, chromosome 9 SNPs (23 SNPs) were informative for 82/84 patients (98%). Only one patient had no SNP informativity for either chromosome; however, this patient had a degraded constitutive control that provided poor results.

2.6.1. SNP-based LOH in experimentally mixed cell population

We sought to determine the accuracy of pyrosequencing assays to detect LOH in mixed cell populations using biological samples. Patients and SNPs were selected to be constitutively heterozygous such that any deviation from the heterozygous range in the experimental mixtures mimics LOH in a mixed cell population. We used the probability model to determine LOH in the experimentally mixed constitutive

control:aneuploid populations described above and did not exclude intermediate LOH calls for this sample set. Each mixture was treated as a sample and we used the LOH algorithm to determine LOH for each sample (see statistical methods). Figure 5 shows two DNA samples genotyped in duplicate for SNP ID 478, with the results of the LOH algorithm binning. The automated algorithm called LOH at levels consistent with expected theoretical mixtures, with LOH being detected in the presence of 75–80% DNA from aneuploid cells.

2.6.2. SNP-based LOH within the CDKN2A locus

The previous STR panel did not include an STR within the CDKN2A/ARF locus, which precluded an accurate assessment of small, intragenic deletions within the CDKN2A/ARF locus using STRs. Therefore, in the 225 fresh/frozen endoscopic biopsy samples, we compared the CDKN2A intragenic SNP LOH calls to LOH calls in SNPs flanking that locus, and to LOH calls made by chromosome 9p STRs. 135 samples had at least one SNP with LOH on chromosome 9p. 31 of 135 events (23%) showed LOH only in SNPs intragenic to the CDKN2A locus, indicating that these were small LOH or deletion events not involving large regions of

chromosome 9p. Of these 31 localized events detected by SNPs, only 12 (39%) were also detected by STRs flanking CDKN2A, indicating that the STR panel likely underestimates the frequency of deletion at the CDKN2A locus on chromosome 9p. 10 out of 135 samples (7%) with LOH detected in both SNPs and STRs flanking CDKN2A along chromosome 9p did not show any LOH or allelic imbalance in intragenic SNPs, possibly due to homozygous deletion. Thus, gene-specific SNP-based LOH alone may not accurately reflect the complex chromosomal instability that is captured by using polymorphisms along the length of the chromosome.

3. Discussion

A main goal of cancer prevention and early detection research is to translate promising biomarkers identified in discovery and early translational research into clinical assays that will be a benefit to patients. Several prospective studies have provided evidence that a chromosomal instability biomarker panel including 17p and 9p LOH and DNA content abnormalities can identify patients with BE at the highest risk of progression to EA [11,16,40–42,52]. In addition, cytology studies in BE and EA using FISH probes to assess copy number changes and digital DNA ploidy analysis have reported a high sensitivity to identify dysplasia and EA from normal squamous esophagus and have shown promise as biomarkers of progression [14,45,46]. Clinical applications of cancer risk stratification biomarkers could greatly facilitate the transformation from treating established disease to a future based on the NIH three “p”s, prediction, prevention and personalized medicine. This manuscript characterizes the analytic performance of a SNP-based platform compared to the validated STR-based LOH biomarker. The single-platform SNP-based LOH panel provides a framework for moving along the translational continuum (<http://www.cancer.gov/trwg/TRWG-definition-and-TR-continuum>) toward late translation including creation of mutually beneficial industry partnerships, and device development. In addition, this single-platform LOH device may be further developed for multi-center studies or to assess entry criteria for randomized cancer prevention trials.

LOH is a common somatic genetic abnormality in solid tumors and many precancerous conditions [8,9, 15,17,22,27,28,33]. The TP53 and CDKN2A/ARF loci are two of the most common targets for inactivation in human cancers associated with LOH [32,43,47,48]. In

addition to 17p and 9p LOH in BE, several longitudinal studies have reported that detection of selected LOH in premalignant tissues can identify patients at increased risk for future progression to cancer, providing a window for early detection and prevention. For example, several longitudinal studies of oral leukoplakia have reported that selected LOH involving 3p, 4q, 8p, 9p, 11q, and 17p can predict progression to cancer [24,29,44]. The majority of these studies have been based on STR microsatellite repeats that are not amenable to emerging high-throughput technologies. The approaches presented in this manuscript may be applied to any targeted chromosomal region and thus transforms a research assay into a clinical platform.

Most biomarker studies of LOH have been generated by amplification of STRs. However, the majority of biorepositories contain only formalin-fixed paraffin-embedded samples, which are not ideal for PCR amplification using STR loci. In addition, STRs have been shown to be problematic with low levels of input DNA [12] or degraded DNA [19]. Sample preparation for STR analysis is a multi-step process with differential amplification efficiencies, multiple sample manipulations, and multiple PCR reactions. The SNP-based protocol developed in this manuscript recommends 0.5 ng DNA (roughly 70 cells) per SNP assay. We have determined that the pyrosequencing technology provides accurate genotypes when using far less input DNA and amplification efficiency with SNPs has improved compared to STRs across samples derived from a variety of sources, such as DNA from fresh-frozen, FFPE, and surgical resection specimens. No allelic drop-off was detected down to as low as 0.015 ng input DNA with different types of input constitutive control DNA (fresh/frozen or FFPE). For DNA from fresh/frozen tissue, the percent passing genotypes remained high down to 0.03 ng (4–5 cell equivalents) (Table 3). Experimental variables inherent in PCR may affect pass rates, but even with reduced pass rates, the genotypes produced are robust, with 99% of genotypes matching across different PCR annealing temperatures and only minor variation in allele ratios of $\leq 3.4\%$ CV. Pyrosequencing provides an internal sequence control to gauge the quality of the sequence across different SNP assays. Each individual SNP assay has built-in quality controls to ensure the accuracy of the genotype. Pyrosequencing is user-friendly, quickly mastered, and does not require extensive technician training. In addition, the pyrosequencing instrument is FDA-approved as a class 1 clinical device.

Pyrosequencing has been used to demonstrate the ability to translate a research-based biomarker into a

clinically compatible SNP-based assay for patient use. Other alternatives to pyrosequencing could include mass spectrometry, Taqman-based allelic discrimination, whole genome SNP arrays, and micro-bead flow-cytometry. Considerations for alternative technologies include flexibility to modify the SNPs being assayed in each protocol, expense and expertise required for devices, and cost. High-throughput whole genome SNP-based array genotyping technologies such as Illumina and Affymetrix genotype indiscriminately, providing results for non-targeted areas of the genome, which poses the problem of obtaining genetic information unrelated to the clinical condition being addressed. Extensive genotyping of SNPs on constitutive DNA may reveal information about susceptibility to diseases, of which a patient would not have provided consent to be evaluated. In addition, high density SNPs in somatic samples likely will identify sub-diagnostic events which could lead to over-diagnosis by clinicians. Micro-bead flow-cytometry (Luminex xMap) has been shown to be comparable to pyrosequencing [7].

Pyrosequencing is a flexible, robust, quantitative SNP-based platform for genotyping. However, it requires informative (heterozygous) SNPs to give LOH results. It has been reported that high-density SNP platforms have been used for LOH without the use of a paired constitutive control [3,31]. However, controlling for constitutive copy number variations and detecting small, somatic genetic events requires a paired, constitutive control. Copy number alterations including copy gain are important alterations in BE and may require a platform which interrogates more SNPs to accurately detect the significant regions of chromosome instability that are predictive of progression or have a high sensitivity and specificity to detect current EA [25, 46] While it is possible to derive absolute copy number at specific loci using pyrosequencing, this technique relies on a diploid genome and determination of copy number is problematic in the setting of chromosomal instability [38,49]. This LOH panel was designed to detect the relative copy number of alleles. SNPs were selected to be both intragenic to CDKN2A and at loci along chromosome 9p and this panel identifies both localized and chromosome arm LOH events, from which homozygous deletion data can be derived. It has been shown that localized homozygous deletions spanning CDKN2A/ARF and other loci are common in BE [39], but the importance of these events as biomarkers of future EA risk is unknown [31].

The LOH panel is intended to be applied to endoscopic clinical samples with results available for pa-

tient care management. This panel is designed such that all SNPs for control and biopsy samples from a single individual can be evaluated simultaneously on a single reaction plate. This universal-plate design also reduces the possibility of cross-patient contamination by allowing LOH assays for single individual to be run on one reaction plate. We have shown that this SNP-based LOH platform replicates previous STR-based LOH biomarkers of cancer risk. The approach used to design and test this SNP panel demonstrates translation from a research-based assay to a clinically compatible platform. The SNP panel can be incorporated as part of cancer risk stratification in multi-center validation studies or to assess entry criteria for intervention studies and clinical trials.

4. Materials and methods

4.1. Samples

155 constitutive control DNA samples extracted from normal gastric biopsies (taken below the diaphragm, approximately 45 cm from the teeth) from 128 patients enrolled in the Seattle Barrett's cohort was used for assay development and optimization and to serve as a paired reference for LOH detection in BE tissue. DNA was derived from different tissue sources including fresh-frozen endoscopic biopsies ($n = 225$), surgical resection samples ($n = 138$), EDTA isolated BE epithelium ($n = 54$) [9] and microdissected BE epithelium from formalin-fixed paraffin-embedded (FFPE) endoscopic biopsy sections ($n = 46$). DNA from fresh frozen biopsies was extracted using either standard phenol/chloroform or the Puregene DNA Isolation Kit (Gentra Systems, Inc. Minneapolis, MN). DNA from FFPE sections was extracted using the Frank method [13]. The Seattle Barrett's Esophagus Study was approved by the Human Subjects Division of the University of Washington in 1983 and renewed annually thereafter with reciprocity from the IRB of the Fred Hutchinson Cancer Research Center (FHCRC) from 1993 to 2001. Since 2001, the study has been approved annually by the IRB of the FHCRC with reciprocity from the Human Subjects Division of the University of Washington. These experiments were undertaken with the understanding and written consent of each subject, and the study conforms with The Code of Ethics of the World Medical Association (Declaration of Helsinki), printed in the British Medical Journal (18 July 1964).

4.2. SNP PCR

The SNP PCR reactions were all performed on genomic DNA from fresh/frozen and FFPE samples with varying input amounts ranging from 0.002 ng up to 6 ng per reaction, depending on the particular experimental variable being tested. PCR conditions consisted of 10X PCR Buffer with no MgCl₂ (Applied Biosystems), 25 mM MgCl₂ (final 1.5 mM) (Applied Biosystems), 10 mM dNTPs (Roche), 1.25 U Amplitaq Gold (Applied Biosystems, cat#N808-0243), and 10 uM of both forward and reverse primers, for a final volume of 25ul. Thermocycling conditions were 94°C for 15 min, followed by 95°C for 15 s, 57°C for 30 s, and 72°C for 15 s, for 45 cycles, and a final extension step at 72°C for 5 minutes. All SNP assays were completed under the same thermocycling conditions. For mixing experiments, constitutive control DNA derived from fundic samples was combined in increasing increments with flow-cytometrically purified aneuploid DNA from the same patient [16,37]. Six test samples from six patients were created by mixing together one constitutive control gastric and one aneuploid sample from the same patient in 0, 10, 25, 50, 75, 90, and 100% constitutive control to aneuploid ratios. 6 ng of each mixture was used per SNP reaction.

4.3. Amplification of STR polymorphisms

Fresh samples were amplified with Primer Extension Preamplification (PEP) according to an established protocol [4]. In mixing experiments, samples were amplified using three STRs on chromosome 9 (D9S1121, D9S925, D9S935) and four STRs on chromosome 17 (D17S1537, D17S786, D17S974, TP53VNTR). For comparisons between SNPs and STRs, we used the same STRs for LOH assessment as in the previously described 10-year prospective study of BE [16]. Fixed samples were amplified from genomic DNA without preamplification. For standard genotyping, a minimum of 50 relative fluorescent units (RFU) was required for defining a peak by the manufacturer (Applied Biosystems Genotyper v3.7). To accurately call LOH in a mixed cell population, STR genotypes for non-informative samples with only one allele were defined as passing if the peak was ≥ 50 RFU. For informative samples with two alleles, passing was defined as both alleles being above ≥ 50 RFU and for informative samples with LOH, at least one allele must be ≥ 125 RFU.

4.4. Template preparation for pyrosequencing

Either the forward or reverse PCR primer was biotinylated on the 5' end to select a single-stranded template for pyrosequencing. 10ul of the PCR product was combined with 2ul of Streptavidin-Sepharose High Performance beads (GE HealthCare, Uppsala), 38ul Binding Buffer (Biotage), and 30ul molecular-grade water for a total of 80ul, and then shaken for 10 minutes on a shaker plate at 1400rpm. The Streptavidin beads bind to the biotinylated primer to ensure single-stranded DNA. The biotinylated PCR product was captured on the probes of the Vacuum Prep Tool (VPT) (Biotage) and remained on the VPT during the subsequent cleaning steps. The samples were washed in 70% EtOH and 0.2 M NaOH for 5 seconds each, and then neutralized with Washing Buffer (Biotage) for 10 seconds. The vacuum pressure was released and the VPT was placed into a 12ul volume of 10 uM sequencing primer and Annealing Buffer (Biotage) in a PSQ HS 96-well plate (Biotage). Liquid retention allowed the ssDNA to be released from the VPT probes. The template was annealed to the primer by heating the plate at 80° for 2 minutes on a heat block with a top iron.

4.5. Pyrosequencing

Sequencing primers were developed using the PSQ Assay Design Software (version 1.0.6) (Biotage). Pyrosequencing was completed on the PSQ HS96A machine at room temperature, according to manufacturer's protocol. Pyrosequencing controls consisted of a No Template Control, which lacked any DNA template and was treated as a regular sample. This control revealed any primer heteroduplexing that occurred between the primers (since a free hydroxyl group on the 3 prime-end can cause background peaks) and it also proved that the PCR reagents were not contaminated with extraneous DNA. Two water blanks were also run, including one water blank that was carried through both the PCR and the sequencing reaction, and one only through the sequencing reaction. In order to validate the assays, each time a new assay was initially tested, a Sequencing Primer Control (SPC) and a Template Only Control (TOC) were run. The SPC checks to see if there was any duplexing with the sequencing primer because the hydroxyl group may cause background noise. The TOC tests whether or not the biotinylated ssDNA self-primers. Passing genotypes were defined by excluding scores of "Failed" in either the Genotype or Allele Quantification analysis setting using the Pyrosequencing

ing Analysis software (PSQ HS96A version 1.2, Biotope) and having at least one nucleotide in the SNP with a peak height ≥ 50 relative luminescence unit (RLU), which is more stringent than previously published minimum criteria of 20 RLU [36] or 30 RLU [51].

4.6. Statistical methods

4.6.1. Heterozygosity estimation and determination of number of SNPs

In order to ensure that the heterozygosity rate calculation was accurately assessed in our samples, we determined informativity in constitutive DNA of 93 patients for all 43 SNPs based on probability models described previously [26]. The mean het rate of the final selected SNPs in this study is 0.331; therefore 12 or more SNPs in a given detection region would ensure with 0.99 probability that there would be at least one or more informative SNPs.

4.6.2. LOH measurement with STR and SNP

Out of the 463 samples from all sample types that were genotyped, 404 samples had at least one informative STR and had genotypes for up to 43 SNPs. 330/404 samples had at least one informative STR and at least one informative SNP on the same chromosome arm. In these 330 samples, we compared the LOH results (yes/no, see below) between the STRs and the SNPs in two ways; regional by-SNP LOH and by-STR locus LOH. First, since each STR can be associated with multiple SNPs, we compared the regional by-SNP LOH results between a specific informative STR, and each single informative SNP associated with that STR. 274/330 samples had at least one informative SNP that matched the corresponding informative STR. Second, by-STR locus LOH includes data from one or more SNPs (up to 10 SNPs per STR) which correlate to the chromosomal location at $\pm 1,500$ basepairs from each of the STRs previously evaluated as part of a biomarker panel. For each informative locus if there is LOH for an STR and LOH in any one of the corresponding SNPs, the LOH results were counted as in agreement.

For each chromosome arm, only 225 fresh-frozen endoscopy samples had the full set of 11 STRs and 34 SNPs (CDKN2A SNPs were not included in the analysis as they did not have a complimentary STR). We compared all STR LOH calls relative to all SNP LOH calls on the same chromosome arm. The LOH calls agreed if any one informative STR and any one informative SNP had LOH, or if none had LOH. On chromosome 9p, 197 out of 225 samples had at least

one informative STR and at least one informative SNP. On chromosome 17p, 221 out of 225 samples had at least one informative STR and one informative SNP.

For the STR-based LOH experiments, multiple constitutive control samples from different human subjects were genotyped at each locus and the ratio of the peak signals of two alleles was calculated. Our data showed that the distribution of allele ratios (log-transformed) of each locus (from different human subjects) satisfied normal distribution well. Therefore, for a given locus (i) it had a locus-specific normal distribution of allele peak ratio values with mean (m_i) and standard deviation (s_i) which was measured from different constitutive control samples. To evaluate the LOH of somatic tissue at a specific locus i for a given patient, the ratio (r_i) of two signal peak values was calculated and log transformed. The p-value of the sample that is normal (LOH no) is estimated as $p_n = P(Z > \frac{r_i - m_i}{s_i}) = 1 - \Phi\left(\frac{r_i - m_i}{s_i}\right)$, where Φ is the cumulative distribution function of normal. The p-value for LOH is estimated by $1 - p_n$. Because multiple loci were evaluated in the study, the p-values were adjusted [5] for multiple comparison with false discovery rate to be $\leq 1\%$. For SNP-based assay data, a SNP has two types of normal distributions, homozygous allele ratio and heterozygous allele ratio. These were calculated similar to the method above for STRs with slightly different methods for p-value calculation. For a given somatic sample, two p-values were calculated based on two normal distributions as above: homozygous SNP value distribution and heterozygous SNP value distribution. The sample p-value $p_s = \text{argmax}(p_{het}, p_{hom})$, where p_{het} and p_{hom} are the p-value calculated from the two distributions (homozygous and heterozygous). The p-values were also adjusted for multiple comparisons with the method described above. An informative sample could be classified as LOH yes or LOH no when $p_s > 0.01$; occasionally, if $p_s < 0.01$, the sample was classified as LOH intermediate (partial LOH or allelic imbalance).

Acknowledgements

This research was funded by NIH P01 CA91955.

References

- [1] A. Ahmadian, B. Gharizadeh, A.C. Gustafsson, F. Sterky, P. Nyren, M. Uhlen and J. Lundeberg, Single-nucleotide polymorphism analysis by pyrosequencing, *Anal Biochem* **280** (2000), 103–110.

- [2] L.A. Anderson, L.J. Murray, S.J. Murphy, D.A. Fitzpatrick, B.T. Johnston, R.G. Watson, P. McCarron and A.T. Gavin, Mortality in Barrett's oesophagus: results from a population based study, *Gut* **52** (2003), 1081–1084.
- [3] G. Assie, T. Laframboise, P. Platzer, J. Bertherat, C.A. Stratakis and C. Eng, SNP arrays in heterogeneous tissue: highly accurate collection of both germline and somatic genetic information from unpaired single tumor samples, *Am J Hum Genet* **82** (2008), 903–915.
- [4] M.T. Barrett, B.J. Reid and G. Joslyn, Genotypic analysis of multiple loci in somatic cells by whole genome amplification, *Nucleic Acids Research* **23** (1995), 3488–3492.
- [5] Y. Benjamini and Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, *J. Roy Stat Soc. Ser B* **57** (1995), 289–300.
- [6] J.D. Birkmeyer, A.E. Siewers, E.V. Finlayson, T.A. Stukel, F.L. Lucas, I. Batista, H.G. Welch and D.E. Wennberg, Hospital volume and surgical mortality in the United States, *N Engl J Med* **346** (2002), 1128–1137.
- [7] S. Bruse, M. Moreau, M. Azaro, R. Zimmerman and L. Brzustowicz, Improvements to bead-based oligonucleotide ligation SNP genotyping assays, *Biotechniques* **45** (2008), 559–571.
- [8] S.S. Chang and J. Califano, Current status of biomarkers in head and neck cancer, *J Surg Oncol* **97** (2008), 640–643.
- [9] I. Cifola, R. Spinelli, L. Beltrame, C. Peano, E. Fasoli, S. Ferrero, S. Bosari, S. Signorini, F. Rocco, R. Perego, V. Proserpio, F. Raimondo, P. Mocarrelli and C. Battaglia, Genome-wide screening of copy number alterations and LOH events in renal cell carcinomas and integration with gene expression profile, *Mol Cancer* **7** (2008), 6.
- [10] J. Dixon, V. Strugala, S.M. Griffin, M.R. Welfare, P.W. Dettmar, A. Allen and J.P. Pearson, Esophageal mucin: an adherent mucus gel barrier is absent in the normal esophagus but present in columnar-lined Barrett's esophagus, *Am J Gastroenterol* **96** (2001), 2575–2583.
- [11] K. Dolan, A.I. Morris, J.R. Gosney, J.K. Field and R. Sutton, Loss of heterozygosity on chromosome 17p predicts neoplastic progression in Barrett's esophagus, *J Gastroenterol Hepatol* **18** (2003), 683–689.
- [12] D.L. Diewer, M.C. Kline, J.W. Redman and J.M. Butler, NIST mixed stain study 3: signal intensity balance in commercial short tandem repeat multiplexes, *Anal Chem* **76** (2004), 6928–6934.
- [13] T.S. Frank, S.M. Svoboda-Newman and E.D. Hsi, Comparison of methods for extracting DNA from formalin-fixed paraffin sections for nonisotopic PCR, *Diagn Mol Pathol* **5** (1996), 220–224.
- [14] E.G. Fritcher, S.M. Brankley, B.R. Kipp, J.S. Voss, M.B. Campion, L.E. Morrison, M.S. Legator, L.S. Lutzke, K.K. Wang, T.J. Sebo and K.C. Halling, A comparison of conventional cytology, DNA ploidy analysis, and fluorescence in situ hybridization for the detection of dysplasia and adenocarcinoma in patients with Barrett's esophagus, *Hum Pathol* **39** (2008), 1128–1135.
- [15] M. Gaasenbeek, K. Howarth, A.J. Rowan, P.A. Gorman, A. Jones, T. Chaplin, Y. Liu, D. Bicknell, E.J. Davison, H. Fiegler, N.P. Carter, R.R. Roylance and I.P. Tomlinson, Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex changes and multiple forms of chromosomal instability in colorectal cancers, *Cancer Res* **66** (2006), 3471–3479.
- [16] P.C. Galipeau, X. Li, P.L. Blount, C.C. Maley, C.A. Sanchez, R.D. Odze, K. Ayub, P.S. Rabinovitch, T.L. Vaughan and B.J. Reid, NSAIDs modulate CDKN2A, TP53, and DNA content risk for future esophageal adenocarcinoma, *PLoS Med* **4** (2007), e67.
- [17] M. Guillaud, L. Zhang, C. Poh, M.P. Rosin and C. Macaulay, Potential use of quantitative tissue phenotype to predict malignant risk for oral premalignant lesions, *Cancer Res* **68** (2008), 3099–3107.
- [18] R.C. Haggitt, Barrett's esophagus, dysplasia, and adenocarcinoma, *Hum Pathol* **25** (1994), 982–993.
- [19] C.R. Hill, M.C. Kline, M.D. Coble and J.M. Butler, Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples, *J Forensic Sci* **53** (2008), 73–80.
- [20] B. Jovov, C.M. Van Itallie, N.J. Shaheen, J.L. Carson, T.M. Gambling, J.M. Anderson and R.C. Orlando, Claudin-18: a dominant tight junction protein in Barrett's esophagus and likely contributor to its acid resistance, *Am J Physiol Gastrointest Liver Physiol* **293** (2007), G1106–1113.
- [21] W.J. Kent, BLAT—the BLAST-like alignment tool, *Genome Res* **12** (2002), 656–664.
- [22] J.N. Kloth, J. Oosting, T. Van Wezel, K. Szuhai, J. Knijnenburg, A. Gorter, G.G. Kenter, G.J. Fleuren and E.S. Jordanova, Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex genetic alterations in cervical cancer, *BMC Genomics* **8** (2007), 53.
- [23] K. Konishi, L. Shen, S. Wang, S.J. Meltzer, N. Harpaz and J.P. Issa, Rare CpG island methylator phenotype in ulcerative colitis-associated neoplasias, *Gastroenterology* **132** (2007), 1254–1260.
- [24] J.J. Lee, W.K. Hong, W.N. Hittelman, L. Mao, R. Lotan, D.M. Shin, S.E. Benner, X.C. Xu, J.S. Lee, V.M. Papadimitrakopoulou, C. Geyer, C. Perez, J.W. Martin, A.K. El-Naggar and S.M. Lippman, Predicting cancer development in oral leukoplakia: ten years of translational research, *Clin Cancer Res* **6** (2000), 1702–1710.
- [25] X. Li, P.C. Galipeau, C.A. Sanchez, P.L. Blount, C.C. Maley, J. Arnaudo, D.A. Peiffer, D. Pokholok, K.L. Gunderson and B.J. Reid, Single nucleotide polymorphism-based genome-wide chromosome copy change, loss of heterozygosity, and aneuploidy in Barrett's esophagus neoplastic progression, *Cancer Prev Res (Phila Pa)* **1** (2008), 413–423.
- [26] X. Li, S.G. Self, P.C. Galipeau, T.G. Paulson and B.J. Reid, Direct inference of SNP heterozygosity rates and resolution of LOH detection, *PLoS Comput Biol* **3** (2007), e244.
- [27] K. Lindblad-Toh, D.M. Tanenbaum, M.J. Daly, E. Winchester, W.O. Lui, A. Villapakkam, S.E. Stanton, C. Larsson, T.J. Hudson, B.E. Johnson, E.S. Lander and M. Meyerson, Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays, *Nat Biotechnol* **18** (2000), 1001–1005.
- [28] K.C. Lo, D. Bailey, T. Burkhardt, P. Gardina, Y. Turpaz and J.K. Cowell, Comprehensive analysis of loss of heterozygosity events in glioblastoma using the 100K SNP mapping arrays and comparison with copy number abnormalities defined by BAC array comparative genomic hybridization, *Genes Chromosomes Cancer* **47** (2008), 221–237.
- [29] L. Mao, J.S. Lee, Y.H. Fan, J.Y. Ro, J.G. Batsakis, S. Lippman, W. Hittelman and W.K. Hong, Frequent microsatellite alterations at chromosomes 9p21 and 3p14 in oral premalignant lesions and their value in cancer risk assessment, *Nature Medicine* **2** (1996), 682–685.
- [30] P. Moayyedi, N. Burch, N. Akhtar-Danesh, S.K. Enaganti, R. Harrison, N.J. Talley and J. Jankowski, Mortality rates in

- patients with Barrett's oesophagus, *Aliment Pharmacol Ther* **27** (2008), 316–320.
- [31] D.J. Nancarrow, H.Y. Handoko, B.M. Smithers, D.C. Gotley, P.A. Drew, D.I. Watson, A.D. Clouston, N.K. Hayward and D.C. Whiteman, Genome-wide copy number analysis in esophageal adenocarcinoma using high-density single-nucleotide polymorphism arrays, *Cancer Res* **68** (2008), 4163–4172.
- [32] M. Olivier, R. Eeles, M. Hollstein, M.A. Khan, C.C. Harris and P. Hainaut, The IARC TP53 database: new online mutation analysis and recommendations to users, *Hum Mutat* **19** (2002), 607–614.
- [33] J. Oosting, E.H. Lips, R. Van Eijk, P.H. Eilers, K. Szuhai, C. Wijmenga, H. Morreau and T. Van Wezel, High-resolution copy number analysis of paraffin-embedded archival tissue using SNP BeadArrays, *Genome Res* **17** (2007), 368–376.
- [34] R.C. Orlando, Mucosal Defense in Barrett's Esophagus, in *Barrett's Esophagus and Esophageal Adenocarcinoma*, Blackwell Publishing, Ltd, Oxford, UK, 2006, pp. 60–72.
- [35] J. Ostrowski, M. Mikula, J. Karczmarzski, T. Rubel, L.S. Wyrwicz, P. Bragoszewski, P. Gaj, M. Dadlez, E. Butruk and J. Regula, Molecular defense mechanisms of Barrett's metaplasia estimated by an integrative genomics, *J Mol Med* **85** (2007), 733–743.
- [36] M.H. Pascual, J.L. Royo, F.J. Martinez-Tello, C. Crespo, A. Salinas, E.F. Herrero, M. Lopez-Garcia, L.M. Real, A. Ruiz and R. Ramirez-Lorca, Exploring allelic imbalance within paraffin-embedded tumor biopsies using pyrosequencing technology, *Clin Chem Lab Med* **44** (2006), 1076–1081.
- [37] T.G. Paulson, P.C. Galipeau and B.J. Reid, Loss of heterozygosity analysis using whole genome amplification, cell sorting, and fluorescence-based PCR, *Genome Research* **9** (1999), 482–491.
- [38] G. Pielberg, A.E. Day, G.S. Plastow and L. Andersson, A sensitive method for detecting variation in copy numbers of duplicated genes, *Genome Res* **13** (2003), 2171–2177.
- [39] E.L. Powell, L.M. Leoni, M.I. Canto, A.A. Forastiere, C.A. Iacobuzio-Donahue, J.S. Wang, A. Maitra and E. Montgomery, Concordant loss of MTAP and p16/CDKN2A expression in gastroesophageal carcinogenesis: evidence of homozygous deletion in esophageal noninvasive precursor lesions and therapeutic implications, *Am J Surg Pathol* **29** (2005), 1497–1504.
- [40] P.S. Rabinovitch, G. Longton, P.L. Blount, D.S. Levine and B.J. Reid, Predictors of progression in Barrett's esophagus III: baseline flow cytometric variables, *American Journal of Gastroenterology* **96** (2001), 3071–3083.
- [41] B.J. Reid, D.S. Levine, G. Longton, P.L. Blount and P.S. Rabinovitch, Predictors of progression to cancer in Barrett's esophagus: baseline histology and flow cytometry identify low- and high-risk patient subsets, *American Journal of Gastroenterology* **95** (2000), 1669–1676.
- [42] B.J. Reid, L.J. Prevo, P.C. Galipeau, C.A. Sanchez, G. Longton, D.S. Levine, P.L. Blount and P.S. Rabinovitch, Predictors of progression in Barrett's esophagus II: baseline 17p (p. 53) loss of heterozygosity identifies a patient subset at increased risk for neoplastic progression, *American Journal of Gastroenterology* **96** (2001), 2839–2848.
- [43] J.W. Rocco and D. Sidransky, p16(MTS-1/CDKN2/INK4a) in cancer progression, *Exp Cell Res* **264** (2001), 42–55.
- [44] M.P. Rosin, X. Cheng, C. Poh, W.L. Lam, Y. Huang, J. Lovas, K. Berean, J.B. Epstein, R. Priddy, N.D. Le and L. Zhang, Use of allelic loss to predict malignant risk for low-grade oral epithelial dysplasia, *Clin Cancer Res* **6** (2000), 357–362.
- [45] A.M. Rygiel, F. Milano, F.J. Ten Kate, J.G. De Groot, M.P. Peppelenbosch, J.J. Bergman and K.K. Krishnadath, Assessment of chromosomal gains as compared to DNA content changes is more useful to detect dysplasia in Barrett's esophagus brush cytology specimens, *Genes Chromosomes Cancer* **47** (2008), 396–404.
- [46] A.M. Rygiel, F. Milano, F.J. Ten Kate, A. Schaap, K.K. Wang, M.P. Peppelenbosch, J.J. Bergman and K.K. Krishnadath, Gains and amplifications of c-myc, EGFR, and 20.q13 loci in the no dysplasia-dysplasia-adenocarcinoma sequence of Barrett's esophagus, *Cancer Epidemiol Biomarkers Prev* **17** (2008), 1380–1385.
- [47] N.E. Sharpless, INK4a/ARF: a multifunctional tumor suppressor locus, *Mutat Res* **576** (2005), 22–38.
- [48] C.J. Sherr, Principles of Tumor Suppression, *Cell* **116** (2004), 235–246.
- [49] E. Soderback, A.L. Zackrisson, B. Lindblom and A. Alderborn, Determination of CYP2D6 gene copy number by pyrosequencing, *Clin Chem* **51** (2005), 522–531.
- [50] M. Sullivan Pepe, R. Etzioni, Z. Feng, J.D. Potter, M.L. Thompson, M. Thornquist, M. Winget and Y. Yasui, Phases of biomarker development for early detection of cancer, *J Natl Cancer Inst* **93** (2001), 1054–1061.
- [51] S.L. Takala, D.L. Smith, O.C. Stine, D. Coulibaly, M.A. Thera, O.K. Doumbo and C.V. Plowe, A high-throughput method for quantifying alleles and haplotypes of the malaria vaccine candidate Plasmodium falciparum merozoite surface protein-1 19 kDa, *Malar J* **5** (2006), 31.
- [52] L. Teodori, W. Gohde, M. Persiani, F. Ferrario, D. Tirindelli Danesi, C. Scarpignato, U. Di Tondo, P. Alo and L. Capurso, DNA/protein flow cytometry as a predictive marker of malignancy in dysplasia-free Barrett's esophagus: thirteen-year follow-up study on a cohort of patients, *Cytometry* **34** (1998), 257–263.
- [53] T. Thomas, K.R. Abrams, J.S. De Caestecker and R.J. Robinson, Meta analysis: Cancer risk in Barrett's oesophagus, *Aliment Pharmacol Ther* **26** (2007), 1465–1477.
- [54] N.A. Tobey, C.M. Argote, X.C. Vanegas, W. Barlow and R.C. Orlando, Electrical parameters and ion species for active transport in human esophageal stratified squamous epithelium and Barrett's specialized columnar epithelium, *Am J Physiol Gastrointest Liver Physiol* **293** (2007), G264–270.
- [55] K.K. Wang and R.E. Sampliner, Updated guidelines 2008 for the diagnosis, surveillance and therapy of Barrett's esophagus, *Am J Gastroenterol* **103** (2008), 788–797.

Copyright of *Cancer Biomarkers* is the property of IOS Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.