Supplementary Material

Oral cancer prediction by noninvasive genetic screening

Jos B. Poell, Leon J. Wils, Arjen Brink, Ralf Dietrich, Christine Krieg, Eunike Velleuer, Ilkay Evren, Elisabeth R. Brouns, Jan G. de Visscher, Elisabeth Bloemena, Bauke Ylstra, Ruud H. Brakenhoff

Table of contents:

Supplementary table S1. Coverage statistics for low-coverage whole genome sequencing of all samples. Sequencing summary statistics for each individual sample, including total reads sequenced, uniquely mapped non-duplicate reads, and used reads for copy number profiling. For coverage the minimum, mean, and maximum sequencing depth, as well as the fraction of bases with coverage equal to or greater than 10X is given. Reads were mapped to build hg19 of the human genome. The table is provided in a separate file.

Supplementary table S2. Coverage statistics for target-enriched sequencing of all samples. Sequencing summary statistics for each individual sample, including total reads sequenced, uniquely mapped non-duplicate reads, and total targeted bases covered. For coverage the minimum, mean, and maximum sequencing depth, as well as the fraction of bases with coverage equal to or greater than 20X, 50X, 100X and 200X is given. Reads were mapped to build hg19 of the human genome. The table is provided in a separate file.

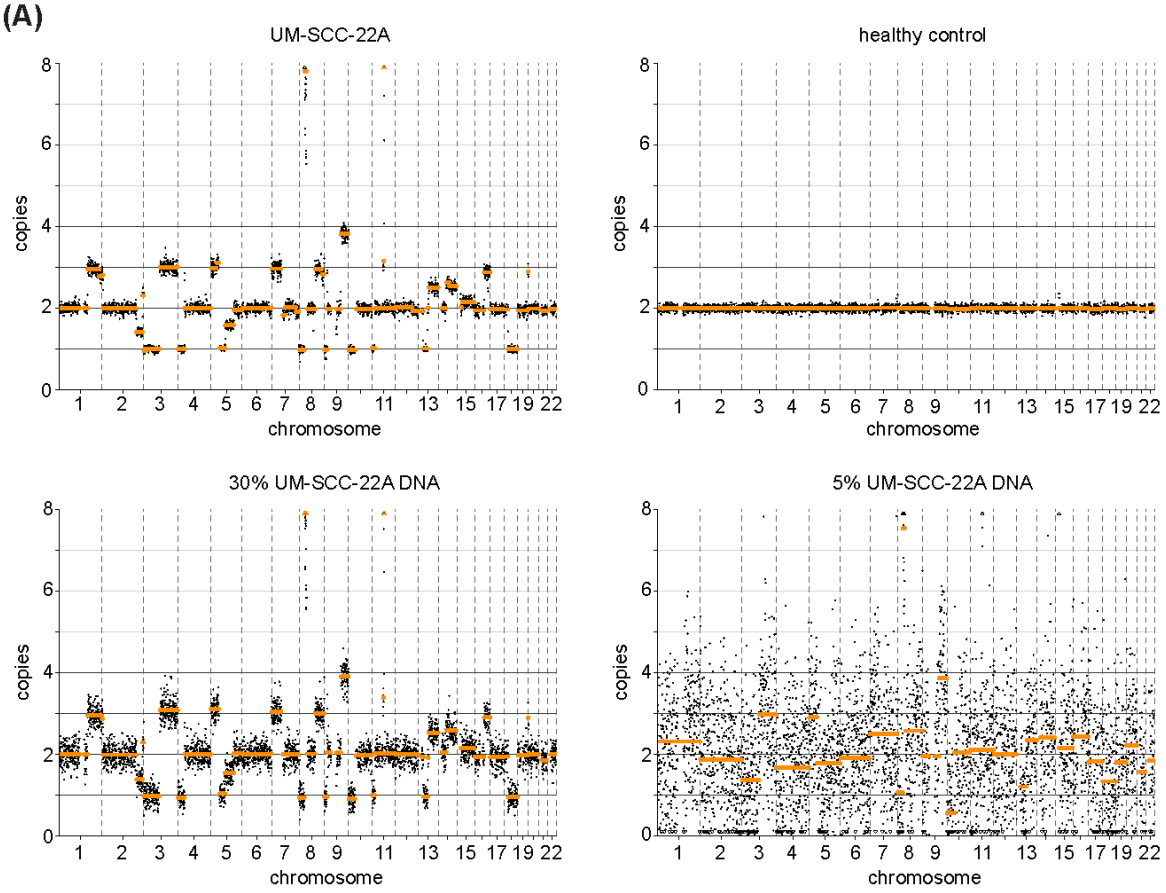
Fanconi anemia cohort	developed OSCC	no OSCC	P-value	test
Total	17	25		
Age (mean ± SD, range)	30.8 ± 11.5, 16–55	29.2 ± 10.5, 16–51	0.65	t-test
Female	8	14	0.75	FET
Male	9	11		
SCT	11	18	0.74	FET
No SCT	6	7		

Supplementary table S3. Distribution of confounding variables in Fanconi Anemia cases and controls. Age, sex, and stem cell transplantation are reported separately for the cases (FA patients who developed OSCC) and the controls (FA patients who did not develop OSCC). Age was compared between cases and controls by unpaired t-test assuming unequal variance. Sex and stem cell transplantation were compared using Fisher's exact test (FET).

Variable	Coefficient	Standard Error	Hazard	95% CI	P-value
Mutations (yes)	1.268	0.511	3.553	1.30–9.68	0.013
Age (years)	0.027	0.034	1.027	0.96–1.10	0.425
SCT (yes)	-0.148	0.678	0.862	0.23–3.26	0.827
Sex (male)	0.291	0.496	1.338	0.51–3.54	0.557

Supplementary table S4. Cox proportional hazards model for development of OSCC. A Cox model for time-to-event (development of OSCC) was fitted for the variables mutations, age, sex, and stem cell transplantation. The coefficient of the variables, which is the natural logarithm of the proportional hazard associated with the variable, is reported, along with its standard deviation, and the hazard with 95% confidence interval and P-value. Hazard applies to the status of the variable as indicated in parentheses for binary variables.

(A)



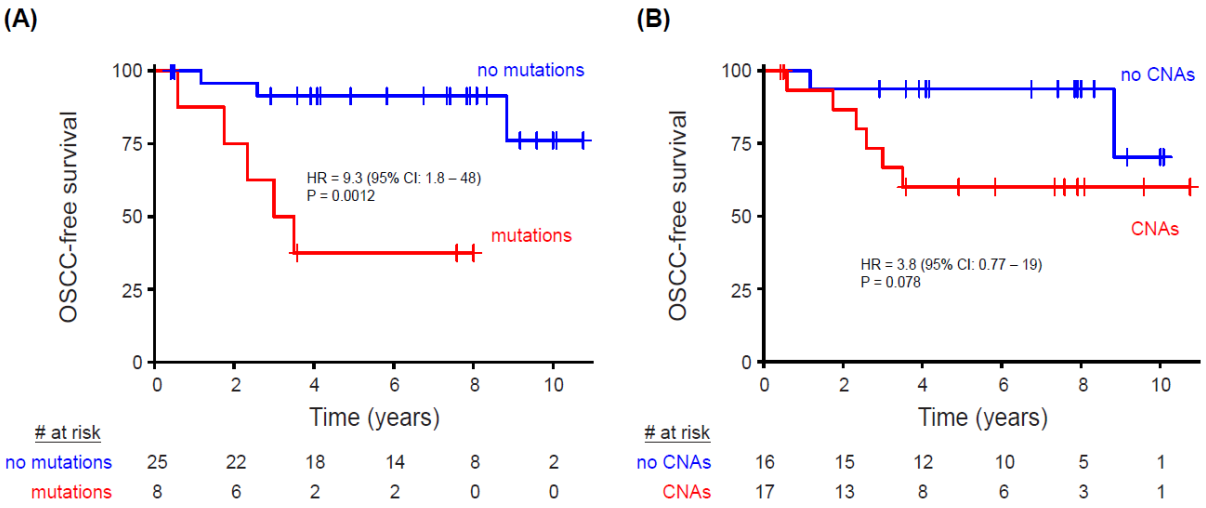
(B)

				variant allele frequency (%)		
Chromosome	**Position**	**Gene**	**Change**	**100% pure**	**30% pure**	**5% pure**
9	139397766	NOTCH1	Glu1679X	99.73	50.28	10.81
17	7576852	TP53	splice donor	47.29	16.02	2.16
17	7578190	TP53	Tyr220Cys	47.95	16.05	2.24

Supplementary figure S1. Detection of CNAs and mutations in cell line DNA diluted with DNA from healthy control cells. DNA of tumor cell line UM-SCC-22A were analyzed separately and in dilutions of 30% UM-SCC-22A DNA and 5% UM-SCC-22A DNA. The respective pure and diluted DNA samples were subjected to low-coverage WGS to determine CNAs (**A**) and targeted deep sequencing to detect mutations in known HNSCC driver genes (**B**). Variant allele frequencies are in accordance with 4 out of 4 mutant NOTCH1 copies in UM-SCC-22A and both copies of TP53 containing a separate mutation.

Supplementary figure S2. Copy number profiles and somatic mutations of all OL brush samples and matching biopsies. Copy number profiles scaled to optimally fit integer copy numbers and somatic mutations with variant allele frequency over 2.5% for brush samples and over 5% for biopsy samples are depicted for all OL samples. Genetic alterations found in matched brushes and biopsies are largely concordant, although the exact composition is often not identical. This figure is provided in a separate PDF-file.

Supplementary figure S3. Copy number profiles and somatic mutations of all brush samples of FA patients. Copy number profiles scaled to optimally fit integer copy numbers and somatic mutations with variant allele frequency over 2.5% are depicted for all brush samples obtained from FA patients. Of all patients, brush samples taken from visually normal-appearing mucosa were analyzed. In addition, visible lesions were analyzed for DE019 and US082. This figure is provided in a separate PDF-file.



Supplementary figure S4. OSCC-free survival of FA patients with minimum follow-up more than 3 months. Time to event from the moment the brush sample was taken until end of follow-up (censored) or detection of OSCC (event) stratified by whether mutations (**A**) or CNAs (**B**) were detected in the brush sample of the respective patients. Patients with diagnosis of OSCC within 3 months after brushing were excluded in this analysis.