Oral Oncology 49 (2013) 211–215

Contents lists available at SciVerse ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma

Edmund A. Mroz [a], James W. Rocco [a,b,*]

[a] Center for Cancer Research and Department of Surgery, Massachusetts General Hospital, Boston, MA, United States
[b] Department of Otolaryngology, Massachusetts Eye and Ear Infirmary and Harvard Medical School, Boston, MA, United States

ARTICLE INFO

Article history:
Received 16 August 2012
Received in revised form 7 September 2012
Accepted 10 September 2012
Available online 15 October 2012

Keywords:
Head and neck cancer
Intratumor heterogeneity
Tumor biomarkers
Next-generation DNA sequencing
Somatic mutations
TP53
Human papillomavirus
Cigarette smoking

SUMMARY

Objectives: Differences among cancer cells within a tumor are important in tumorigenesis and treatment resistance, yet no measure of intratumor heterogeneity is suitable for routine application. We developed a quantitative measure of intratumor genetic heterogeneity, based on differences among mutated loci in the mutant-allele fractions determined by next-generation sequencing (NGS) of tumor DNA. We then evaluated the application of this measure to head and neck squamous cell carcinoma (HNSCC).
Materials and methods: We analyzed published electronically available NGS results for 74 HNSCC. For each tumor we calculated mutant-allele tumor heterogeneity (MATH) as the ratio of the width to the center of its distribution of mutant-allele fractions among tumor-specific mutated loci.
Results: Intratumor heterogeneity assessed by MATH was higher in three poor-outcome classes of HNSCC: tumors with disruptive mutations in the TP53 gene (versus wild-type TP53 or non-disruptive mutations), tumors negative versus positive for human papillomavirus (even when restricted to tumors having wild-type TP53), and HPV-negative tumors from smokers with more pack-years of cigarette exposure (with TP53 status taken into account).
Conclusion: The relation of this type of intratumor heterogeneity to HNSCC outcome classes supports its further evaluation as a prognostic biomarker. As NGS of tumor DNA becomes widespread in clinical research and practice, MATH should provide a simple, quantitative, and clinically practical biomarker to help evaluate relations of intratumor genetic heterogeneity to outcome in any type of cancer.

Introduction

Differences among cancer cells within a tumor are important in disease progression, metastasis, and treatment resistance, with heterogeneous tumors more likely to have developed a subpopulation that is therapy-resistant or metastasis-prone.[1,2] A measure of this intratumor heterogeneity might provide clinically significant information. Unfortunately, the techniques used to establish the importance of intratumor heterogeneity – such as examining intratumor distribution of pre-identified markers,[3–5] extensive tumor dissection,[3,4,6,7] isolating and analyzing individual nuclei,[3,5,8] and ultradeep sequencing of mutations[9] – are difficult to translate from research studies to the clinic.

We propose a way to use results of next-generation sequencing (NGS), expected to be applied soon in clinical oncology, to obtain a measure of intratumor genomic heterogeneity. Genomically distinct subpopulations of cells in a tumor lead to differences among mutated loci in terms of the fraction of sequence reads that show a mutant allele.[9] The distribution of mutant-allele fractions among loci thus provides a straightforward measure of one type of intratumor heterogeneity, called mutant-allele tumor heterogeneity (MATH). MATH represents a consequence of multiple cell populations in a tumor, while avoiding the practical and theoretical difficulties of trying to identify and enumerate them directly. Using published NGS results on head and neck squamous cell carcinoma (HNSCC),[10] we show that MATH is high in each of three poor-outcome classifications of HNSCC.

Materials and methods

Clinical data and NGS exome-sequencing results (approximately 1% of the genome, at 150-fold mean sequence coverage) for 74 HNSCC, and data on genomic copy-number alterations (CNA) for 55 of these, were imported from Supplementary Tables 6, 10 and 11 of Stransky et al.[10] into R.[11] Each tumor's MATH value was calculated from the median absolute deviation (MAD) and the median of its mutant-allele fractions at tumor-specific mutated loci:

* Corresponding author at: Jackson 904G, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, United States. Tel.: +1 617 726 5251; fax: +1 617 726 8623.
E-mail address: jrocco@partners.org (J.W. Rocco).

1368-8375/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.oraloncology.2012.09.007

$$\text{MATH} = 100 * \text{MAD}/\text{median}.$$

Calculation of MAD followed the default in R, with values scaled by a constant factor (1.4826) so that the expected MAD of a sample from a normal distribution equals the standard deviation.

Following Poeta et al.[12] we scored TP53 mutations whose predicted protein products were truncated or had altered charge or polarity of an amino-acid residue in the L2 or L3 binding domains as disruptive. Splice-site mutations were also scored as disruptive. HPV status was based on reported PCR results.[10] Significance was taken at $p < 0.05$ in 2-sided tests.

Results

Calculating MATH from NGS results

MATH is the ratio of the width to the center of the distribution of mutant-allele fractions among tumor-specific mutated loci. The basic idea is illustrated in Fig. 1A for an idealized situation with heterozygous loci, no CNA, and no normal tissue. A heterogeneous tumor will tend to have a wider distribution of mutant-allele fractions among loci, centered at a lower fraction, than a homogeneous tumor. The width of the distribution captures diversity among loci arising from different cell populations. Taking the ratio of the width to the center provides a first-order correction for the presence of normal tissue in the tumor, because the multiplicative factor correcting total cell numbers for normal-cell numbers appears in both the numerator and the denominator. Robust measures of the width and the center, the MAD and median, are used to minimize the influence of outlier loci (e.g., low-fraction loci with less precise values, and high-fraction loci common to most cell populations or with CNA favoring the mutant allele). Supplementary Text provides details on the relation of distribution width to patterns of mutation sharing among cell populations and CNA, the use of robust measures, and how the ratio of MAD to median corrects for normal tissue.

MATH ranged from 19 to 55 (dimensionless units) among 74 HNSCC. Distributions of mutant-allele fractions and the corresponding MATH values are shown for three cases in Fig. 1B. CNA data available for a subset of 55 samples showed little influence of CNA on MATH. More than 90% of mutated loci had genomic copy numbers within 0.5 $\log_2$ units of normal, so that MATH values were similar whether calculated from all mutated loci (as in data presented here), only from loci having low CNA, or from mutant-allele fractions corrected for local CNA (Supplementary Text; Supplementary Figs. S1, S2; Supplementary Table S1).

MATH and mutation rate

Although intratumor heterogeneity and mutation rate are different concepts, it was possible that tumors with high mutation rates would simply have greater intratumor heterogeneity. This was not the case. MATH was not related to overall mutation rate, conventionally expressed as the number of mutated loci per MB of sequenced DNA,[10] as illustrated by the three examples in Fig. 1B and by a plot of MATH versus the number of mutated loci (Fig. 1C). MATH thus represents a different aspect of tumor biology than the mutation rate itself.

As MATH is calculated from single-molecule DNA sequence reads, the precision of MATH values depends on sampling of loci and of mutant versus reference alleles. We estimated the associated standard deviation (SD) of the MATH value for each tumor by bootstrap resampling from all sequence reads at mutated loci (median, 12,600 reads per tumor). At the median of 92 mutated loci per sample, the SD was 4 units; the SD decreased with the square root of the number of mutated loci (Fig. 1D). Thus the precision of determining MATH is better in samples with higher mutation rates, even though MATH values themselves are not related to mutation rate.

High intratumor heterogeneity in three poor-outcome classifications of HNSCC

Using MATH as a measure of intratumor heterogeneity, we examined the relationship of heterogeneity to three variables whose importance has been established clinically in HNSCC: disruptive TP53 mutations,[12,13] human papillomavirus (HPV) status,[14,15] and increasing exposure to cigarette smoke.[16]

First, we examined the relation of intratumor heterogeneity to mutations in the TP53 gene. The role of the p53 protein as guardian of the genome[17] suggests a general hypothesis that TP53 mutations would lead to increased intratumor heterogeneity. Nevertheless, the multiple roles of p53 and the different functional consequences of different types of mutations[18] mean that not all TP53 mutations might influence heterogeneity. In HNSCC, only a subset of TP53 mutations, called "disruptive,"[12] is related to worse outcome, while patients with non-disruptive mutations have similar outcomes as those with wild-type TP53.[12,13] We thus tested the hypothesis that disruptive rather than non-disruptive TP53 mutations are associated with greater intratumor genetic heterogeneity in HNSCC.

Consistent with this hypothesis, disruptive mutations in TP53 were specifically related to higher MATH values. MATH was higher in HNSCC having disruptive TP53 mutations than in those with non-disruptive mutations ($p = 0.038$) or wild-type sequence ($p = 0.008$), but did not differ between tumors having non-disruptive mutations and wild-type TP53 ($p = 0.93$; Fig. 2).

Second, we examined the relation of intratumor heterogeneity to HPV status. HPV infection contributes to a large and growing number of HNSCC cases,[15] with HPV-positive HNSCC usually discovered at a younger patient age, having a lower overall mutation rate,[10] exhibiting fewer large-scale genomic copy-number changes,[19] and typically having better outcomes than HPV-negative tumors.[14,15] With the p53 and p16-pRb tumor suppressor pathways inactivated by HPV gene products,[20] somatic mutations silencing these pathways are not needed before development of invasive tumors. The less extensive history of mutation and selection in HPV-positive than in HPV-negative tumors at time of presentation leads to the hypothesis that HPV-positive tumors would have less intratumor heterogeneity.

Consistent with this hypothesis, MATH was lower in HPV-positive than in HPV-negative cases ($p = 0.011$; Fig. 3, left versus center). To rule out a simple explanation based on the lack of TP53 mutations in these HPV-positive tumors,[10] we also restricted analysis to tumors having wild-type TP53. There was still a significant difference in MATH between HPV-positive and HPV-negative cases having wild-type TP53 ($p = 0.047$; Fig. 3, center versus right).

Third, we examined the relation of intratumor heterogeneity to cigarette use. Exposure to cigarette smoke not only is a mutagenic risk factor for HNSCC but also is associated with worse outcome following therapy, with each pack-year of cigarette use increasing the hazard ratio for relapse or death by 1%.[16] Increased exposure to mutagens as measured by pack-years, generating clonal subpopulations of cells thought to underlie progression and field cancerization in HPV-negative HNSCC,[21] would be predicted to lead to greater intratumor heterogeneity.

Consistent with this hypothesis, MATH was significantly associated with cigarette pack-years, among smokers having HPV-negative tumors, when TP53 mutation status was taken into account (Table 1). Each 10 extra pack-years was associated with an increase of 1.1 MATH units in a tumor.

Note that for classifications of HNSCC by each of these clinical variables, the class having the highest heterogeneity as assessed