Abbreviations

AUC	Area under the curve
ECOG	Eastern Cooperative Oncology Group
EORTC-QLQ-C30	European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30
NPV	Negative predictive value
PPV	Positive predictive value
PRO	Patient-reported outcome
ROC	Receiver operating characteristic
SCNS-SF34	Supportive Care Needs Survey-Short Form-34

Introduction

The use of patient-reported outcome (PRO) measures in clinical practice for individual patient management involves having a patient complete a questionnaire about his/her functioning and well-being and providing that patient's scores to his/her clinician to inform care and management [1, 2]. The procedure is analogous to laboratory tests that inform the clinician about the patient's health—the difference being that PROs are based on scores from patient-reported questionnaires rather than values from chemical or microscopic analyses. The use of PROs for individual patient management has been consistently shown to improve clinician–patient communication [3–6]. It has also been shown to improve detection of problems [6–9], affect management [5], and improve patient outcomes, such as symptom control, health-related quality-of-life, and functioning [3, 10, 11].

Although we have demonstrated that PROs can effectively identify the issues that are bothering patients the most [12], an ongoing challenge to the use of PROs in clinical practice is determining which scores require a clinician's attention. That is, after patients complete the PRO questionnaire, their responses are scored and a score report is generated. However, for clinicians reviewing the scores, it is not intuitive which scores represent a problem that should motivate action. Various methods have been applied to assist with score interpretation, including providing the mean score for the general population for comparison [3] or highlighting scores using the lowest quartile from the general population as a cutoff [13]. However, these methods do not actually reflect whether a score represents an unmet need from the perspective of the patient, which would require a clinician's attention.

To address this issue, in a previous study, we used the Supportive Care Needs Survey-Short Form (SCNS-SF34) to determine cutoff scores on the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire-Core 30 (QLQ-C30) that identify unmet needs [14]. We demonstrated that QLQ-C30 scores can discriminate between patients with and without unmet needs; however, the study was conducted in a limited sample ($n = 117$) of breast, prostate, and lung cancer patients from a single institution. The present analysis was undertaken to attempt to replicate the findings using a new and larger sample.

Patients and methods

Research design and data source

The objective of this study was to test the replicability of the QLQ-C30 cutoff scores from our previous study. To address this objective, we conducted a secondary analysis of data originally collected in the validation study of the Japanese version of the Supportive Care Needs Survey-Short Form (SCNS-SF34-J). The methods of this Japanese study have been reported previously [15]. Briefly, ambulatory breast cancer patients were recruited from the Oncology, Immunology and Surgery outpatient clinic of Nagoya City University Hospital. Inclusion criteria included diagnosis of breast cancer, age at least 20 years, awareness of cancer diagnosis, and Eastern Cooperative Oncology Group (ECOG) performance status of 0–3. Exclusion criteria were severe mental or cognitive disorders or inability to understand Japanese. Participants were selected at random using a list of visits and a random number table to limit the number of patients enrolled each day.

After providing written consent, subjects completed a paper survey that included the SCNS-SF34-J (validated in the parent study [15]) and the Japanese version of the EORTC-QLQ-C30 (described below). In addition to these PRO questionnaires, the survey included basic sociodemographic questions. Patients were instructed to return the completed survey to the clinic the following day, and follow-up by telephone was used to clarify inadequate answers. The attending physician provided ECOG performance status, and information on cancer stage and treatments was abstracted from the patients' medical records.

The SCNS-SF34 was originally developed by investigators in Australia to identify unmet needs cancer patients have in five domains: physical and daily living, psychological, patient care and support, health system and information, and sexual [16, 17]. The 34-item questionnaire uses five response options: 1 = not applicable, 2 = satisfied, 3 = low unmet need, 4 = moderate unmet need, and 5 = high unmet need and a recall period of the "last month." To calculate domain scores, we averaged the

scores of the items within the domain; thus, domain scores >2.0 reflected some level of unmet need.

The QLQ-C30 [18] is a cancer health-related quality-of-life questionnaire that has been widely used in clinical trials and investigations using PROs for individual patient management [3, 6, 11, 19]. It includes five function domains (physical, emotional, social, role, and cognitive), eight symptoms (fatigue, pain, nausea/vomiting, constipation, diarrhea, insomnia, dyspnea, and appetite loss), as well as global health/quality-of-life and financial impact. Subjects respond on a four-point scale from "not at all" to "very much" for most items. Most items use a "past week" recall period. Raw scores are linearly converted to a 0–100 scale with higher scores reflecting higher levels of function and higher levels of symptom burden. The Japanese version of the QLQ-C30 has been validated previously [20].

The Japanese study was approved by the Institutional Review Board and Ethics Committee of Nagoya City University Graduate School of Medical Sciences [15]. A de-identified dataset was provided to the Johns Hopkins investigators for this analysis, which was exempted for review by the Johns Hopkins School of Medicine Institutional Review Board.

Analyses

The data were analyzed using the methods applied in the original study using the SCNS-SF34 to identify cutoff scores on the QLQ-C30 that represent unmet need [14]. First, we dichotomized the SCNS-SF34 item and domain scores into no unmet need (scores ≤ 2.0) versus some unmet need (scores > 2.0). We then tested the ability of QLQ-C30 domain scores to discriminate between patients with and without an unmet need using the SCNS-SF34 domains and items we tested in our previous analysis (see Table 1 for a summary of the SCNS-SF34 items/domains tested for each QLQ-C30 domain). Variables for the discriminant analysis were selected to correspond as closely as possible to the content of the QLQ-C30 domains. In some cases, the content was quite similar (e.g., pain on the QLQ-C30 and pain on the SCNS-SF34). For a few QLQ-C30 domains, there was no SCNS-SF34 item or domain with similar content. In these cases we used a generic SCNS-SF34 item such as "feeling unwell a lot of the time."

The discriminative ability of each QLQ-C30 domain score was summarized using the area under the receiver operating characteristic (ROC) curve (AUC). The AUC summarizes the ability of QLQ-C30 scores to discriminate between patients with and without a reported unmet need. Higher AUCs indicate better discriminative ability. For the domains with AUC ≥ 0.70, we then calculated the sensitivity and specificity, as well as the positive and negative predictive values, associated with various QLQ-C30 cutoff scores. We used a threshold of AUC ≥ 0.70 because Hosmer and Lemeshow suggest that values below 0.70 represent poor discrimination, between 0.70 and 0.80 represent acceptable discrimination, and above 0.80 represent excellent discrimination [21]. It was also the standard used for our previous analysis [14]. We hypothesized that compared to our original analysis, (1) the same QLQ-C30 domains would have AUC ≥ 0.70, (2) the same SCNS-SF34 items would be best discriminated by the QLQ-C30 and thus provide the highest AUC, and (3) the sensitivity and specificity of our original cutoff scores would be supported. Analyses were performed using statistical freeware R version 2.15.1.

Results

The sample has been described previously [15]. Briefly, from a pool of 420 potential participants, 12 were excluded due to declining participation ($n = 7$), cognitive deficits ($n = 2$), advanced disease ($n = 1$), and failure to respond after consenting ($n = 2$). The study sample included 408 subjects with a mean age of 56 years, 100 % female, 76 % married, and 45 % employed full- or part-time. The ECOG performance status was 0 for 90 % of the sample; the clinical stage was I or II for 71 %; 93 % had received surgery, 44 % chemotherapy, and 39 % radiation; and the median time from diagnosis was 701 days (range 11–17,915 days). Complete data were available for all 408 subjects, with the exception of one participant who was missing a single SCNS-SF34 item. That observation was excluded from analyses that required that item.

Table 1 shows which SCNS-SF34 items/domains were used to evaluate the discriminative ability for each QLQ-C30 domain, as well as the resulting AUCs both from our original analysis [14] and from this replication analysis. The AUCs were largely similar between studies. As hypothesized, the same six QLQ-C30 domains with AUCs ≥ 0.70 in the original analysis had AUCs ≥ 0.70 in the replication sample. Further, the SCNS-SF34 item that was best discriminated by the QLQ-C30 with the highest AUC in the original analysis also had the highest AUC in the replication sample. The following QLQ-C30 domain–SCNS-SF34 item pairings were used: physical function–work around the home (AUC = 0.74), role function–work around the home (AUC = 0.70), emotional function–feelings of sadness (AUC = 0.75), pain–pain (AUC = 0.74), fatigue–lack of energy/tiredness (AUC = 0.75), and global health/QOL–feeling unwell a lot of the time (AUC = 0.76).

Using these pairings, we evaluated the sensitivity, specificity, and predictive value of various cutoff scores on the QLQ-C30 (Table 2). Again, the results were largely similar between the original analysis and this replication

Table 1 Hypothesized relationship between QLQ-C30 and SCNS-SF34 domains and resulting areas under the curve (AUC): original and replication analysis

QLQ-C30 Domain	SCNS-SF34 Domain/Item(s)	AUC	
		Original Analysis [14]	Replication Analysis
Hypothesized AUC≥0.70			
Physical Function	Physical & daily living needs (overall score and individual items)	0.69–0.81	0.69–0.74
Role Function	Work around the home Not being able to do the things you used to	0.71–0.73	0.70–0.70
Emotional Function	Psychological needs (overall score and individual items)	0.56–0.74	0.61–0.75
Pain	Pain	0.78	0.74
Fatigue	Lack of energy/tiredness	0.74	0.75
Global Health /QOL	Feeling unwell a lot of the time	0.73	0.76
Hypothesized AUC <0.70			
Social Function	Not being able to do the things you used to	0.64	0.68
Sleep	Lack of energy/tiredness Feeling unwell a lot of the time Being given information about aspects of managing your illness and side effects at home	0.41–0.51	0.39–0.55
Cognitive Function	Feeling unwell a lot of the time Being given information about aspects of managing your illness and side effects at home	0.54–0.60	0.53–0.63
Nausea/vomiting		0.19–0.36	0.22–0.27
Dyspnea		0.37–0.48	0.32–0.48
Appetite Loss		0.47–0.49	0.32–0.49
Constipation		0.31–0.37	0.32–0.40
Diarrhea		0.34–0.34	0.18–0.21

sample. Examples of cutoff scores (sensitivity, specificity) from the replication sample are as follows: physical function <90 (0.85, 0.65); role function <90 (0.85, 0.62); emotional function <90 (0.84, 0.60); global health/QOL < 70 (0.86, 0.56); pain >10 (0.93, 0.54); and fatigue >30 (0.86, 0.62). Thus, each domain had at least one cutoff score with sensitivity ≥0.84 and specificity ≥0.54. This means that patients who reported unmet needs in a domain were identified correctly at least 84 % of the time and that patients who reported no unmet needs in a domain were identified correctly at least 54 % of the time using these cutoffs. In general, the negative predictive values (NPVs) associated with these cutoffs were higher than the positive predictive values (PPVs), with the NPVs ranging from 0.86 to 0.94 and PPVs ranging from 0.33 to 0.58. This means that if a patient was identified by the cutoff as not having an unmet need in a domain, 86–94 % of the time they did not report an unmet need and that if a patient was identified by the cutoff as having an unmet need, 33–58 % of the time they actually did report an unmet need. While we describe these cutoff scores for illustrative purposes, the specific cutoff scores used in a given application should be determined based on the relative importance of sensitivity and specificity.

Discussion

This analysis was undertaken to test the generalizability of the findings from our previous study which evaluated the ability of different cutoff scores on the QLQ-C30 to identify patients with an unmet need in a given domain. Such cutoff scores facilitate the interpretation of PROs used clinically for individual patient management by helping clinicians determine which scores deserve further attention. Currently, there are few guides available to help clinicians determine which PRO scores represent a problem. For example, in PatientViewpoint, the PRO webtool used at Johns Hopkins [13, 22], we highlight in yellow QLQ-C30 domain scores representing the lowest quartile based on published general population norms [23] as an indication to the clinician reviewing the report that the patient may be having a problem in this area. However, these cutoff scores using distributions of the data are not empirically based on whether the score is likely to represent a problem from the patient's perspective. For example, the results from this analysis suggest that domain scores <90 on role or emotional function likely represent a patient-reported unmet need. However, at our institution, we are currently using cutoff scores <66.7 for these two

Table 2 Sensitivity and specificity of example cutoff scores: original and replication analysis

QLQ-C30 Domain	SCNS-SF34 Item	Cutoff	Cohort	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
Physical Function	Work around the home	80	Original [14]	0.65	0.83	0.55	0.89
			Replication	0.40	0.92	0.63	0.82
		90	Original [14]	0.85	0.58	0.39	0.92
			Replication	0.85	0.65	0.45	0.93
Role Function	Work around the home	80	Original [14]	0.69	0.79	0.50	0.89
			Replication	0.69	0.79	0.52	0.88
		90	Original [14]	0.85	0.69	0.46	.94
			Replication	0.85	0.62	0.43	0.93
Emotional Function	Feelings of sadness	90	Original [14]	0.89	0.53	0.48	0.91
			Replication	0.84	0.60	0.58	0.86
		100	Original [14]	0.94	0.35	0.41	0.93
			Replication	0.92	0.42	0.51	0.89
Global Health/QOL	Feeling unwell a lot of the time	70	Original [14]	0.71	0.69	0.52	0.84
			Replication	0.86	0.56	0.33	0.94
		80	Original [14]	0.89	0.58	0.50	0.91
			Replication	0.89	0.45	0.29	0.94
Pain	Pain	20	Original [14]	0.66	0.84	0.64	0.85
			Replication	0.70	0.81	0.62	0.86
		10	Original [14]	0.91	0.66	0.54	0.95
			Replication	0.93	0.54	0.47	0.94
Fatigue	Lack of energy/ tiredness	30	Original [14]	0.77	0.71	0.73	0.75
			Replication	0.86	0.62	0.54	0.90
		20	Original [14]	0.91	0.55	0.68	0.86
			Replication	0.97	0.42	0.46	0.97

domains, based on the population distribution of scores. This means that our current cutoffs are missing patients with unmet needs with scores between 67 and 90. Based on the results of this analysis, we will explore changing the cutoffs to those presented here to highlight QLQ-C30 scores for the clinician's attention.

Our findings should be interpreted in the context of the study's strengths and limitations. First, the approach of using the SCNS-SF34 to identify QLQ-C30 cutoff scores only works well for the six QLQ-C30 domains where there is content overlap between the SCNS-SF34 and QLQ-C30. For the domains without a corresponding SCNS-SF34 item to use for comparison, we do not have indicators of appropriate cutoffs. Future research could address this issue by using items similar in format to the SCNS-SF34 but covering the content of the relevant QLQ-C30 domains for which no data are currently available. Also, the SCNS-SF34 uses a recall period of the "past month," whereas the QLQ-C30 generally uses a recall period of the "past week." Ideally, the comparison between scores would be made with questionnaires that use the same recall period. The study design used in both the current sample and the original analysis was cross-sectional, so while absolute cutoff scores can be identified, important changes in scores are not addressed. Research from longitudinal studies using both the QLQ-C30 and SCNS-SF34 could explore changes in scores representing an unmet need.

Notably, this validation sample used QLQ-C30 and SCNS-SF34 data collected using the Japanese versions of the questionnaires. That we found such similarity between our original analysis and the current sample, despite differences in language and culture, suggests that these findings are robust. While the Japanese study provided a new sample to test our original cutoffs, and almost four times as many patients, only breast cancer patients were enrolled in the Japanese study, whereas our original analysis included three different cancer types (breast, prostate, and lung). Also, the Japanese sample included women with a wide range of time since diagnosis (11–17,915 days). The symptom burden for women who had completed treatment years previously may be lower than for women in active treatment. Nevertheless, given the substantial concordance between this replication sample and our original sample, we believe there is adequate evidence to support implementing these cutoffs in Patient-Viewpoint and other applications of the QLQ-C30 being used in clinical practice.

The next important step will be to evaluate whether clinicians and patients find these cutoffs helpful. A key

consideration is which cutoff to use. We presented several example cutoff scores for illustrative purposes here, but the cutoff scores appropriate for a specific application depend on the relative importance between sensitivity and specificity. That is, the more likely a cutoff score is to identify patients with unmet needs (true positives), the more likely it will also identify patients without an unmet need (false positives). Thus, it is important to consider the implications of false positives versus false negatives.

In general, the use of PROs for individual patient management involves helping the clinician identify problems the patient may be experiencing and facilitating a focused discussion of PRO topics that might otherwise go unaddressed. This is essentially a screening function. We therefore expect follow-up of a "positive" score based on the cutoff to involve the clinician simply asking the patient about the issue and determining whether there is something that can and should be done to address any unmet needs. Given that this requires a minimal effort, it may be appropriate to favor high sensitivity over high specificity. However, it is also important to avoid alert fatigue, a phenomenon that leads to clinician inattention to potential problems and resistance to the tools in general. In addition, if the cutoff scores were to be applied by, for example, generating an automatic page to the clinician, then false positives would be much more problematic. Another issue is how to address PRO scores representing an unmet need. In previous research, we developed a range of suggestions for how to address issues identified by PRO questionnaires [24]. However, it is important to consider resource and reimbursement limitations for certain services (e.g., psychosocial services, home care), as well as their effectiveness, before implementing them as part of care pathways. Consideration of how these cutoff scores will be applied in practice will help determine the appropriate compromise between sensitivity and specificity.

In summary, this analysis was conducted to replicate our original analysis to determine whether specific cutoff scores effectively identify patients with unmet needs. For the QLQ-C30 domains with appropriate SCNS-SF34 content matches, our findings from the original analysis were largely supported. This suggests that these cutoff scores could be applied in practice, with an evaluation of their effectiveness from the clinician and patient perspectives. Specifically, it will be important to see how clinicians actually respond when presented with information from PROs using these (or other appropriate) cutoffs and whether the information helps increase clinicians' awareness of unmet needs. Further research is also needed to identify cutoff scores for QLQ-C30 domains without SCNS-SF34 content matches, as well as to identify changes in scores that represent unmet need. In the meantime, the results for these six domains provide critical guidance to clinicians interpreting PRO reports on which scores require their attention.

Acknowledgments This analysis was supported by the American Cancer Society (MRSG-08-011-01-CPPB). The original data collection was supported in part by Grants-in-Aid for Cancer Research and the Third Term Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health, Labour and Welfare, Japan. Drs. Snyder and Carducci are members of the Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins (P30CA006973). The funding sources had no role in study design, data collection, analysis, interpretation, writing, or decision to submit the manuscript for publication.

Conflict of interest The authors report no conflict of interest.

References

1. Snyder, C. F., & Aaronson, N. K. (2009). Use of patient-reported outcomes in clinical practice. *The Lancet, 374*, 369–370.
2. Greenhalgh, J. (2009). The applications of PROs in clinical practice: What are they, do they work, and why? *Quality of Life Research, 18*, 115–123.
3. Velikova, G., Booth, L., Smith, A. B., et al. (2004). Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomized controlled trial. *Journal of Clinical Oncology, 22*, 714–724.
4. Berry, D. L., Blumenstein, B. A., Halpenny, B., et al. (2011). Enhancing patient-provider communication with the electronic self-report assessment for cancer: A randomized trial. *Journal of Clinical Oncology, 29*, 1029–1035.
5. Santana, M. J., Feeny, D., Johnson, J. A., et al. (2010). Assessing the use of health-related quality of life measures in the routine clinical care of lung-transplant patients. *Quality of Life Research, 19*, 371–379.
6. Detmar, S. B., Muller, M. J., Schornagel, J. H., Wever, L. D. V., & Aaronson, N. K. (2002). Health-related quality-of-life assessments and patient-physician communication. A randomized clinical trial. *Journal of the American Medical Association, 288*, 3027–3034.
7. Greenhalgh, J., & Meadows, K. (1999). The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: A literature review. *Journal of Evaluation in Clinical Practice, 5*, 401–416.
8. Marshall, S., Haywood, K., & Fitzpatrick, R. (2006). Impact of patient-reported outcome measures on routine practice: A structured review. *Journal of Evaluation in Clinical Practice, 12*, 559–568.
9. Haywood, K., Marshall, S., & Fitzpatrick, R. (2006). Patient participation in the consultation process: A structured review of intervention strategies. *Patient Education and Counseling, 63*, 12–23.
10. Cleeland, C. S., Wang, X. S., Shi, Q., et al. (2011). Automated symptom alerts reduce postoperative symptom severity after cancer surgery: A randomized controlled clinical trial. *Journal of Clinical Oncology, 29*, 994–1000.
11. McLachlan, S.-A., Allenby, A., Matthews, J., et al. (2001). Randomized trial of coordinated psychosocial interventions based on patient self-assessments versus standard care to improve the psychosocial functioning of patients with cancer. *Journal of Clinical Oncology, 19*, 4117–4125.
12. Snyder, C. F., Blackford, A. L., Aaronson, N. K., et al. (2011). Can patient-reported outcome measures identify cancer patients' most bothersome issues? *Journal of Clinical Oncology, 29*, 1216–1220.

13. Snyder, C. F., Blackford, A. L., Wolff, A. C., et al. (2012). Feasibility and value of PatientViewpoint: a web system for patient-reported outcomes assessment in clinical practice. *Psycho-Oncology,*. doi:10.1002/pon.3087.
14. Snyder, C. F., Blackford, A. L., Brahmer, J. R., et al. (2010). Needs assessments can identify scores on HRQOL questionnaires that represent problems for patients: an illustration with the Supportive Care Needs Survey and the QLQ-C30. *Quality of Life Research, 19*, 837–845.
15. Okuyama, T., Akechi, T., Yamashita, H., et al. (2009). Reliability and validity of the Japanese version of the short-form supportive care needs survey questionnaire (SCNS-SF34-J). *Psycho-Oncology, 18*, 1003–1010.
16. Bonevski, B., Sanson-Fisher, R. W., Girgis, A., et al. (2000). Evaluation of an instrument to assess the needs of patients with cancer. *Cancer, 88*, 217–225.
17. Sanson-Fisher, R., Girgis, A., Boyes, A., et al. (2000). The unmet supportive care needs of patients with cancer. *Cancer, 88*, 226–237.
18. Aaronson, N. K., Ahmedzai, S., Bergman, B., et al. (1993). The European organization for research and treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute, 85*, 365–376.
19. Velikova, G., Brown, J. M., Smith, A. B., & Selby, P. J. (2002). Computer-based quality of life questionnaires may contribute to doctor-patient interactions in oncology. *British Journal of Cancer, 86*, 51–59.
20. Kobayashi, K., Takeda, F., Teramukai, S., et al. (1998). A cross-validation of the European organization for research and treatment of cancer QLQ-C30 (EORTC QLQ-C30) for Japanese with lung cancer. *European Journal of Cancer, 34*, 810–815.
21. Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). Chichester, New York: Wiley.
22. Snyder, C. F., Jensen, R., Courtin, S. O., Wu, A. W., & Website for Outpatient QOL Assessment Research Network. (2009). PatientViewpoint: A website for patient-reported outcomes assessment. *Quality of Life Research, 18*, 793–800.
23. Fayers, P. M., Weeden, S., Curran, D., & on behalf of the EORTC Quality of Life Study Group. (1998). *EORTC QLQ-C30 Reference Values*. Brussels: EORTC (ISBN: 2-930064-11-0).
24. Hughes, E. F., Wu, A. W., Carducci, M. A., & Snyder, C. F. (2012). What can I do? Recommendations for responding to issues identified by patient-reported outcomes assessments used in clinical practice. *Journal of Supportive Oncology, 10*, 143–148.

革新的癌ワクチン，H/K-HELPの開発

——ショートペプチドからヘルパー/キラーロングペプチドへの移行

Helper/killer-hybrid epitope long peptide(H/K-HELP)

大竹淳矢　増子和尚　角田健太郎　富樫裕二　北村秀光　西村孝司(写真)

Junya Ohtake[1], Kazuhisa Masuko[1], Kentaro Sumida[1], Yuji Togashi[3], Hidemitsu Kitamura[1] and Takashi Nishimura[1,2]

北海道大学遺伝子病制御研究所免疫制御分野[1]，同ROICE' 健康バイオ研究部門[2]，株式会社バイオイミュランス[3]

◎本稿では，ヘルパーエピトープとキラーエピトープを化学的に結合させた人工癌抗原ロングペプチド(helper/killer hybrid epitope long peptide：H/K-HELP)癌ワクチン開発に至る癌免疫の基盤研究と，その成果の臨床研究への応用，そしてなぜH/K-HELPロングペプチドワクチンが，従来のショートペプチドに比べ有効であるかを解き明かす基盤研究の成果について概説したい．

Key word：癌ワクチン治療，H/K-HELP癌ワクチン，ヘルパーT細胞，Th1細胞，CTL

1991年のテリー・ブーン博士らによる癌抗原の発見によって，癌に対する特異的免疫誘導が可能であることが示された[1]．アミノ酸8〜9個からなるクラスⅠ結合性癌抗原キラーペプチドを用いた癌ワクチン治療の臨床研究は，一時は無効とされたが，最近では制癌剤との併用により，あるいは数種のペプチドを混合したマルチペプチドを用いて，癌組織の縮小は認められない場合が多いものの，癌特異的CTLが弱いながらも誘導され，癌患者の生存日数が大幅に延長されることが示されている[2]．さらに，最近ではクラスⅡ結合性癌抗原ヘルパーペプチドの同定もなされ，ヘルパーT細胞とキラーT細胞の両者を活性化できるsynthetic long peptide(SLP)の混合ワクチンがオランダのMelief らによって開発され，HPVで誘発されるヒト外陰部上皮異形成の治療効果があることが示された[3]．

著者らは30数年に及ぶ基盤的癌免疫研究から，より有効な癌免疫治療を開発するためには，①担癌生体の免疫抑制性癌エスケープ機構の解明と，②宿主免疫抑制を打破するためのヘルパーT細胞，とくに癌特異的Th1細胞の活性化が重要であることを提唱してきた[4-6]．最近，①に関してはあらたな分子メカニズムや免疫抑制性細胞群が明らかにされ，さらには癌治療抵抗性を担う癌幹細胞の存在も明らかになってきている．また，②に関しては癌抗原ヘルパーエピトープの同定により癌特異的Th1細胞の誘導が可能となり，さらに，ヘルパーエピトープとキラーエピトープを化学的に結合させた人工癌抗原ロングペプチド(helper/killer hybrid epitope long peptide：H/K-HELP)も開発され，臨床研究においてTh1依存的免疫の誘導効果や癌消失効果も証明されている[7]．

担癌生体における免疫抑制・癌エスケープ機構

癌患者末梢血リンパ球は健常人のそれに比べ，異常にT細胞応答が低下している．これは癌が増殖とともに宿主の免疫応答を抑制し，癌が増殖しやすい場を形成するためと考えられる．従来は，①癌細胞におけるMHCの消失，②癌細胞あるいは免疫担当細胞による免疫抑制因子(TGF-βやIL-10)の産生などが免疫逃避のおもなメカニズムとして報告されてきた．しかし最近は，担癌生体の癌局所において異常に集積する$CD4^+Foxp3^+$制御性T細胞(Treg)，$CD11b^+Gr\text{-}1^+$未

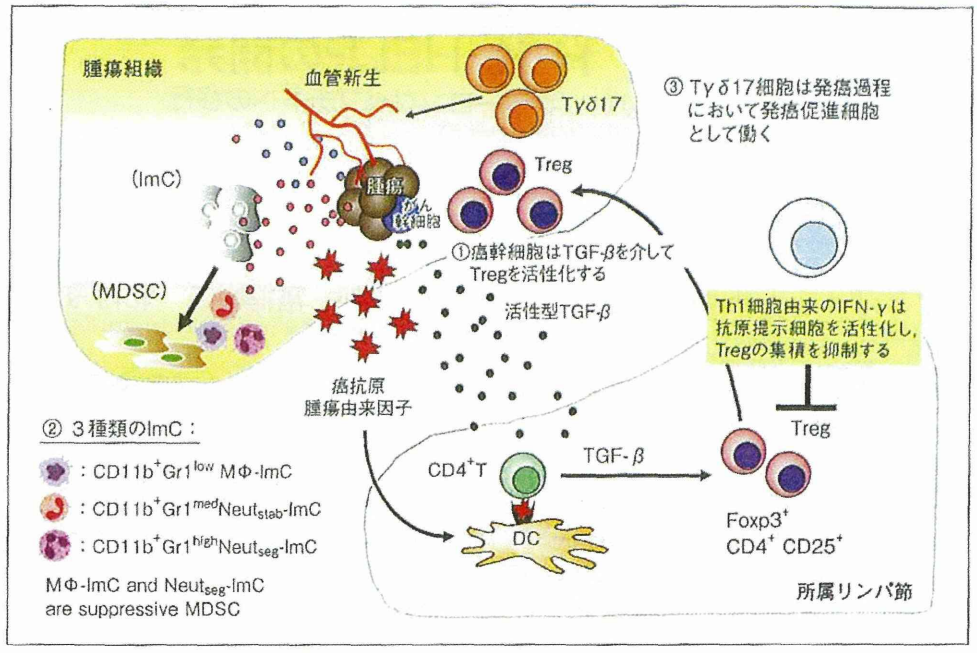

図 1　担癌生体内における免疫逃避機構

腫瘍は自身の生存や増殖に有利な微小環境を形成し，異常な増殖を可能にしている．著者らは腫瘍内微小環境において，①制御性 T 細胞(Treg)，②骨髄由来免疫抑制細胞(MDSC)，③IL-17 産生 γδT 細胞(Tγδ17)が誘導され，それぞれ癌免疫監視機構からの逃避や腫瘍増殖の維持に関与していることを見出している．

熟ミエロイド細胞(immature myeloid cells：ImC)，ImC 由来のサプレッサーマクロファージなど(総称して myeloid-derived suppressor cells：MDSC)による免疫抑制が注目を浴びている[8,9]．また，著者らは IL-17 産生細胞が腫瘍血管形成を促進し，癌の増殖を促進する protumor 細胞であることも明確にしている[10](図 1)．誌面の都合上，詳細については他の総説[11]を参考にしていただき，ここでは著者らが見出した免疫抑制・癌エスケープに関するあらたな知見についてのみ概説する．

1. 癌幹細胞とTreg

$CD133^+$癌幹細胞が非癌幹細胞よりも TGF-β 産生能が高く，$CD133^+$癌幹細胞を接種した担癌マウスの所属リンパ節では $CD133^-$非癌幹細胞担癌マウスに比べ，$CD4^+Foxp3^+$Treg が TGF-β 依存的に多く誘導されていることが明らかとなった．最近，なぜ癌幹細胞が活性型 TGF-β を高産生するのかに関するエピジェネティック分子メカニズムもほぼ解明され，「$CD133^+$癌幹細胞は非癌幹細胞に比べ，活性型 TGF-β を高産生し，Treg を誘導することによって，$CD8^+$CTL を介した癌免疫監視機構から逃避している」ことを明らかにしている(図 1-①)．

2. 担癌生体における ミエロイド由来免疫抑制性細胞(MDSC)

担癌マウス脾内では，①$CD11b^+Gr\text{-}1^{low}$：F4/80$^+$マクロファージ(MΦ-ImC)，②$CD11b^+Gr\text{-}1^{mid}$：好中球桿状核球($Neut_{stab}$-ImC)，③$CD11b^+Gr\text{-}1^{high}$：好中球分葉核球($Neut_{seg}$-ImC)，の 3 つのサブセットが異常増殖する．Gabrilovichi らは，3 つのサブセットすべてを含む細胞群を ImC suppressor あるいは MDSC と定義している[9]．しかし著者らは，免疫抑制活性を示すのは MΦ-ImC のみで，他の $Neut_{stab}$-ImC，や $Neut_{seg}$-ImC は免疫抑制を示さず，脾内に集積したこれら ImC 群は腫瘍内に移住後に，IL-6 や TGF-β などの腫瘍微小環境サイトカインの影響で免疫抑制性 MDSC (MΦ-ImC と $Neut_{seg}$-ImC)へと形質変換し，非常に強い免疫抑制を示すことを証明している．したがって，真の MDSC は MΦ-ImC と $Neut_{seg}$-ImC のみと考えられる．著者らは抗 IL-6R 抗体と

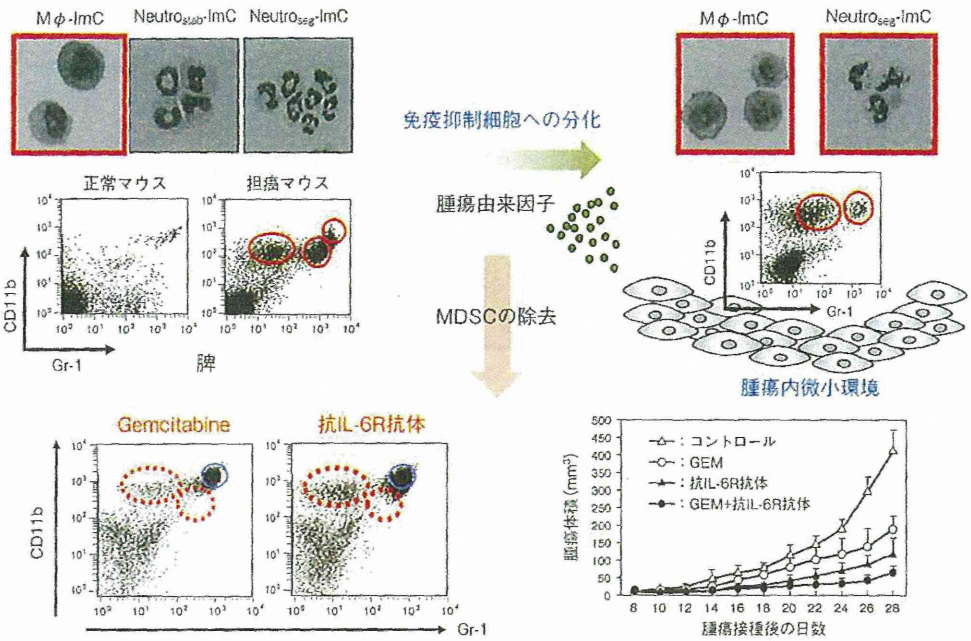

図 2　担癌生体におけるミエロイド由来免疫抑制細胞

担癌生体内で異常増殖する CD11b$^+$Gr-1$^+$細胞は免疫抑制細胞として MDSC と総称されているが，脾内では免疫抑制能を示さない Neut$_{seg}$-ImC，Neut$_{stab}$-ImC，免疫抑制能を有する MΦ-ImC の 3 種のサブセットが存在する．これらの細胞は腫瘍内環境に曝されることによって強い免疫抑制能を有した Neut$_{seg}$-ImC，MΦ-ImC へと分化する．また，抗 IL-6R 抗体と gemcitabine(GEM)の併用投与により MDSC が除去され，抗腫瘍免疫を増強し，腫瘍増殖を抑制することができる．

GEM の併用投与により，免疫抑制を担う MΦ-ImC と Neut$_{stab}$-ImC が選択的に除去され，担癌生体の T 細胞応答が増強され，より効果的な抗腫瘍免疫が誘導されることを証明した(図 1-②，図 2)[12]．

3. 癌組織に浸潤したIL-17産生γδT細胞はprotumor細胞である

BALB/c マウス腫瘍組織内に浸潤する IL-17 産生細胞群を調べてみると，CD4$^+$T 細胞や CD8$^+$T 細胞ではなく，γδT 細胞がおもな産生細胞であることがわかった．このγδT 細胞は皮膚常在性γδT 細胞ではなく，全身循環しているγδT 細胞であり，腫瘍組織へ浸潤後に，TCR を介した抗原刺激，NKG2D を介した補助刺激，あるいは癌微小環境で産生される IL-6，TGF-β，IL-23 の刺激を受けTγδ17 に分化することが明らかにされた[10]．IL-17 KO マウスではメチルコラントレン誘発 caricinoma の形成が腫瘍血管の新生とともに抑えられるので，IL-17 は腫瘍血管新生促進を介して発癌を促進する因子と考えられる(図 1-③)．また著者らは，IL-17 の扁平上皮癌誘発における促進効果は遺伝子支配されており，Th2 マウスである BALB/c では発癌初期における Tγδ17 細胞の浸潤と並行して癌幹細胞様細胞の異常増殖，慢性炎症の遷延化ならびに carcinoma の形成が認められるのに対して，Th1 マウスの C57BL/6 マウスにおいては，Tγδ17 細胞よりも，Th17 細胞が浸潤し，皮膚の肥厚・炎症は早期に沈静化し，carcinoma は形成されず，fibrosarcoma のみが誘発されることを見出している(図 3)．炎症の質の違いと異なる組織癌の発生に関する興味深い知見である．

Th17，Tc17，Tγδ17 などの IL-17 産生細胞が発癌や癌の増殖過程において，protumor，antitumor のいずれとして作用を示すのかという論争が続いたが，著者らはそれに対する解答を Tc17 の研究で最近明らかにした[13]．すなわち，IL-17 産生細胞は通常の微小環境では protumor として機能し，炎症や血管新生に関与するが，Th1 環境下ではエピジェネティックな可塑的変化を遂げ，キ

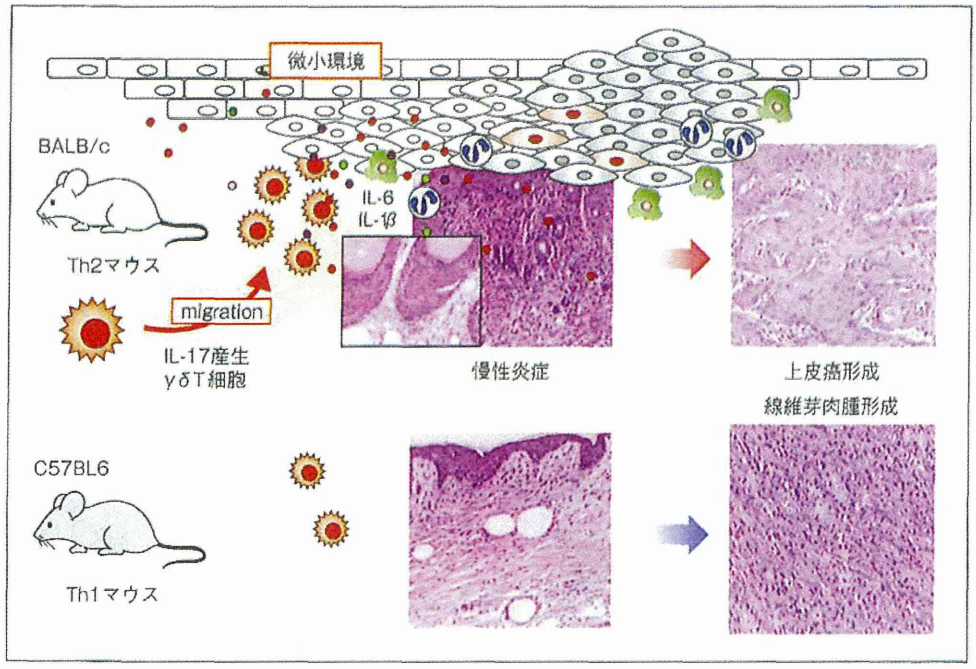

図 3　IL-17産生γδT細胞は発癌プロモーター細胞(protumor cell)として働く
発癌過程の局所組織ではIL-17産生γδT細胞が浸潤し，慢性炎症を誘起することで上皮癌の発生の促進に寄与している．IL-17を介した炎症応答制御は強い遺伝子支配を受け，Th2マウスであるBALB/cマウスではIL-17産生γδT細胞が誘導され，上皮癌が発生するのに対し，Th1マウスであるC57BL/6マウスではIL-17産生γδT細胞の誘導は弱く，線維芽肉腫の発生が認められた．

ラー活性を有するIL-17/IFN-γ double producing Tc17/IFN-γへと変換し細胞傷害性能力を獲得，antitumorエフェクター細胞として抗腫瘍免疫にも関与すると思われる[14]．

Th1主導免疫の導入による担癌生体免疫抑制の打破

担癌生体においては，癌の増殖とともに癌のエスケープを助ける強い免疫抑制が誘導される．したがって，癌患者に癌ワクチン療法を施行する場合にはいかにしてこの負の免疫監視機構を克服し，癌特異的キラーT細胞(Tc；CTL)を誘導するかが重要な課題となる．「腫瘍から浸潤リンパ球を濃縮して培養すると，リンパ球は存在するにもかかわらず，数日で死滅し，混在する癌細胞のみが増殖してくる．しかし，ここに一滴のIL-2を加えると癌を殺しながら増殖するキラー(LAK)が誘導される」．著者が30数年前の学生時代に出くわしたセレンディピティーである(図4)．その日から，IL-2を産生するヘルパーT細胞が腫瘍局所に存在すれば，担癌生体の免疫抑制は克服できるに違いないと考え，①樹状細胞(DC)による癌抗原のプロセシング，②ヘルパーT細胞(Th)による抗原認識，活性化，そして，③癌特異的キラーT細胞(Tc；CTL)の強い活性化誘導までの，自然免疫から獲得免疫までの一連の反応がTh1主導免疫を活性化できるタイプ1免疫依存的に進行することが重要であることを提唱してきた[4-6]．

IL-12，α-GalCel+IL-12，CpG，OK-432，Poly-I：C，Th1細胞などがタイプ1免疫を誘導する手段としてあげられ，TLRを介したリガンドとしてはCpGがもっとも優れたTh1免疫誘導アジュバントと考えられる．事実，CpGをアジュバントとして癌抗原ワクチンを担癌動物に摂取することにより非常に強い抗腫瘍免疫が誘導される．しかし，Tregの免疫抑制を打破できるという点においてはCpGよりも癌特異的Th1細胞でTh1主導免疫を導入する方が優れている[6]．Th1細胞と癌抗原を腫瘍近傍に投与することによって担癌マウス生体に癌特異的CTLを効率よく誘導でき，