Supplementary Online Content

Grossman D, Okwundu N, Bartlett EK, et al. Prognostic gene expression profiling in cutaneous melanoma: identifying the knowledge gaps and assessing the clinical benefit. *JAMA Dermatol*. Published online July 29, 2020. doi:10.1001/jamadermatol.2020.1729

This supplementary material has been provided by the authors to give readers additional information about their work.

eTable 1. Summary of survey responses of melanoma experts.

1st round survey: Potential applications of GEP testing	**Perceived clinical impact**		
	Low	**Medium**	**High**
Predict SLNB positivity in patients with T1a tumors	20 (29%)	15 (22%)	33 (49%)
Predict SLNB positivity in patients with T1b tumors	18 (27%)	17 (25%)	32 (48%)
Identify patients with node-negative T1b tumors who should undergo more intense clinical and/or imaging surveillance	19 (29%)	15 (23%)	32 (48%)
Identify stage II patients who should undergo more intensive clinical/imaging surveillance	8 (12%)	18 (27%)	40 (61%)
Identify stage II patients who would benefit from systemic adjuvant therapy	7 (11%)	8 (12%)	51 (77%)
Identify stage IIIA patients who could undergo less intensive clinical/radiologic surveillance	7 (11%)	17 (26%)	42 (64%)
Identify stage IIIA patients who would benefit from systemic adjuvant therapy	7 (11%)	10 (15%)	49 (74%)
1st round survey: General clinical trial considerations	**Disagree**	**Neutral**	**Agree**
Should employ retrospective sample sets prior to pursuing prospective clinical trials to determine best uses of GEP testing	26 (39%)	10 (15%)	30 (46%)
Future trials should incorporate multiple testing platforms	4 (6%)	17 (26%)	44 (68%)
2nd round survey: GEP test performance in stage I patients (for prediction of distant mets)	**100%**	**95%**	**90%**	**80%**	**70%**
Minimum acceptable PPV	4 (9%)	5 (12%)	13 (30%)	11 (26%)	10 (23%)
Minimum acceptable NPV	5 (12%)	18 (42%)	9 (21%)	4 (9%)	7 (16%)
2nd round survey: Favored trial objectives regarding GEP testing and SLNB	**Worth pursuing**	**Not worth pursuing**
Can GEP testing reliably predict SLNB positivity?	35 (81%)	8 (19%)
Is GEP testing more accurate than SLNB in predicting recurrence or metastasis?	31 (72%)	12 (28%)
2nd round survey: Favored trial objectives regarding GEP testing and SLNB		
GEP testing can complement SLNB and AJCC staging to improve prognostication	16 (37%)	

GEP testing can replace SLNB as a staging test	1 (2%)				
GEP testing can inform management strategies such as surveillance and adjuvant therapy better than SLNB	26 (61%)				
2nd round survey: Highest priority GEP testing trial objective					
GEP testing can replace SLNB as a staging test	3 (7%)				
GEP testing can complement SLNB and AJCC staging to improve prognostication	9 (23%)				
GEP testing can identify patients who could be spared imaging surveillance and adjuvant therapy	12 (30%)				
GEP testing can identify patients who could benefit from imaging surveillance and adjuvant therapy	16 (40%)				
2nd round survey: Favored trial objective regarding GEP testing, imaging surveillance, and adjuvant therapy					
GEP testing can identify those who may benefit from these interventions	4 (10%)				
GEP testing can identify those who may safely avoid these interventions	0 (0%)				
Both are equally important	37 (90%)				
2nd round survey: Clinical trial considerations regarding GEP testing, imaging surveillance, and adjuvant therapy	**Disagree**		**Agree**		
Prior to a trial, it is important first to determine whether early detection of asymptomatic disease by surveillance imaging improves outcomes in patients treated with adjuvant therapy.	17 (41%)		24 (59%)		
GEP testing of primary tumors from stage II patients from completed placebo-controlled trials of anti-PD1 may be sufficient to determine whether GEP testing could be used to determine which stage II patients may benefit from adjuvant therapy.	13 (32%)		28 (68%)		
2nd round survey: Which melanoma stage does GEP testing have the greatest potential to impact patient management?	**IA/B**	**IIA**	**IIB/C**	**IIIA**	**IV**
	9 (23%)	11 (28%)	16 (40%)	3 (7%)	1 (2%)

AJCC, American Joint Committee on Cancer; GEP, gene expression profile; NPV, negative predictive value; PPV, positive predictive value; SLNB, sentinel lymph node biopsy.

[a] A total of 195 melanoma experts were emailed two separate surveys. There were 78 (40%) respondents to one or both surveys. There were 73 respondents to the 1st round survey and 44 respondents to the 2nd round survey. There were 28 respondents to both surveys. This likely reflects some survey fatigue by the 1st round respondents and response to email reminders by some 2nd round respondents who did not respond to the 1st round survey. The respondents were very representative of the MPWG, with the following breakdown: 1st round: 53% dermatologists, 21% medical oncologists, and 10% surgical oncologists; 2nd round: 52% dermatologists, 25% medical oncologists, and 11% surgical oncologists. A subset of the authors formulated the questions for the first-round survey, discussed results, and formulated questions for the second-round survey during two conference phone calls. For each survey question, respondents were also invited to enter free-text comments. Some individuals did not answer some of the questions in each survey.

eTable 2. Hypotheses and potential trial designs for evaluation of GEP testing.

Design type	Total N (per/arm)	Alpha (type-1 error)	Power	3-year OS control arm	3-year OS experimental arm
Hypothesis: Using GEP test results instead of SLNB improves survival					
Superiority	808 (404/arm)	5% (two-sided)	80%	85%	90%
Superiority	1082 (541/arm)	5% (two-sided)	90%	85%	90%
Hypothesis: Survival is not worse after using GEP test results instead of SNLB					
Non-inferiority	7136 (3518/arm)	5% (one-sided)	90%	85%	Not worse than 83%
Non-inferiority	9018 (4509/arm)	5% (one-sided)	95%	85%	Not worse than 83%
Hypothesis: For stage II-IIIA patients with a high-risk GEP test result, survival is better if they receive adjuvant therapy compared to not receiving adjuvant therapy					
Superiority	1196 (598/arm)	5% (two-sided)	80%	90%	93.4%
Superiority	1602 (801/arm)	5% (two-sided)	90%	90%	93.4%
Hypothesis: For stage IIIB (or higher) patients with a low-risk GEP test result, survival is not worse if they do not receive adjuvant therapy compared to receiving adjuvant therapy					
Non-inferiority	4482 (2241/arm)	5% (one-sided)	90%	75%	Not worse than 71.9%
Non-inferiority	5644 (2822/arm)	5% (one-sided)	95%	75%	Not worse than 71.9%

OS, overall survival.
Superiority trials evaluate whether an experimental arm has better outcome(s) compared to a randomized control arm; and analyses are typically intention-to-treat, with patients analyzed based on assigned therapy. Non-inferiority trials evaluate whether an outcome(s) in an experimental arm is at least as good as a control arm; analyses are typically per protocol and require more patients. Sample size requirements obtained using https://stattools.crab.org, assuming 1:1 randomization, 3 years of accrual and an additional 4 years of follow-up, and exponential survival. Estimated sample size requirements will vary depending on the design, specified statistical tests, alpha level, power, expected event-rate, accrual rate, and follow-up time. In a trial with overall survival as the primary outcome, a Cox regression model may be used for the primary analysis and a one-sided 95% confidence interval for the hazard ratio may be calculated. A trial may be powered so that the upper limit of the confidence interval is not larger than 1.05 (indicating that the experimental arm may have no more than 5% greater hazard of death compared to the control arm). The alpha level for non-inferiority trials is often taken to be 5-10% (one-sided) and power is often 90-95%, reflecting the different objectives compared to superiority trials. The alpha and power are selected based on the specific hypothesis of the trial, and some non-inferiority trials can use smaller alpha (1% or smaller). In the non-inferiority calculations above, sample size was calculated assuming that the control and experimental arms

had the same 3-year OS as that listed for the control arm, while the values of the 3-year OS in the experimental arm denote the non- inferiority limit.