Comparison of Two Quantitative Image Analysis Systems for Breast Cancer Immunohistochemistry

Reid H Phillips, BS, Sue DiRenzo, BS, Charalambos C Solomides, MD

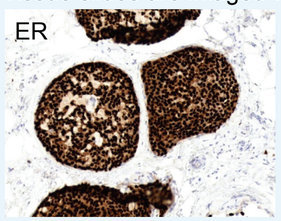
Department of Pathology Anatomy and Cell Biology, Jefferson Medical College of Thomas Jefferson University, Philadelphia, PA

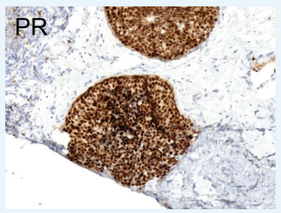
ABSTRACT

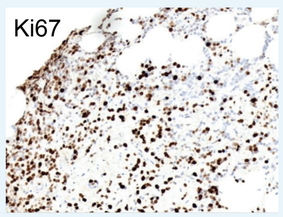
Automated image analysis systems for breast cancer immunohistochemistry promise efficiency and reliability in the quantification of therapy targets such as the estrogen receptor (ER) or human epidermal growth factor receptor (Her2). Thomas Jefferson University Hospital owns two such systems, the Aperio ScanScope AT (Leica Biosystems) and Ventana iScan Coreo (Roche). A comparison study was performed to determine if choice of system affects target quantification and subsequent clinical tumor classification. Tumor expressions of ER, progesterone receptor (PR), proliferation marker Ki67, and Her2 were quantified with both systems for tissue samples from twenty breast cancer patients. Positive tumor classification was based on percent positivity values of ER >1%, PR >1%, Ki67 >10%, and Her2 score of 3+. Percent agreement for tumor classification was ER 100 (95% CI 83.2-100), PR 95 (75.1-99.9), Ki67 90 (68.3-98.8) and Her2 100 (83.2-100). While agreement was high large variation was observed in percent positivity scores (ER, PR, Ki67).

INTRODUCTION

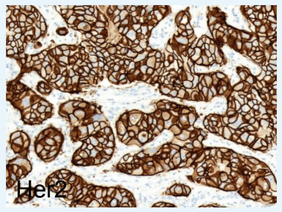
Targeted breast cancer therapies include endocrine therapies such as those targeting the estrogen receptor (ER) and immunotherapies such as those targeting the human epidermal growth factor receptor (Her2)[1,2]. Choice of therapy depends on target expression as quantified by immunohistochemistry (IHC). At Thomas Jefferson University Hospital (TJUH) all breast cancer tissue is stained for ER, PR, Ki67 and Her2. Tissue slides are imaged with specialized digital scanners and image analysis software is used to quantify target expression. Such software allows for an analysis of many more cells than could be reasonably achieved by a pathologist using a microscope. It can be used to improve inter-pathologist reliability and may be a solution to the problem of inconsistency across institutions[1,3].







ER, PR and Ki67 are quantified with a percent positivity, that is the percent of cells staining positive for the target divided by the total number of cells analyzed. Tumors are ER positive if the ER percent positivity is greater than 1%. Patients with ER positive tumors are likely to be treated with an endocrine therapy such as tamoxifen, fulvestrant, or an aromatase inhibitor[1]. Similarly, PR positive tumors are tumors with more than 1% PR positive cells[1]. ER negative, PR positive tumors are uncommon but worth identifying as they may also respond to the above endocrine

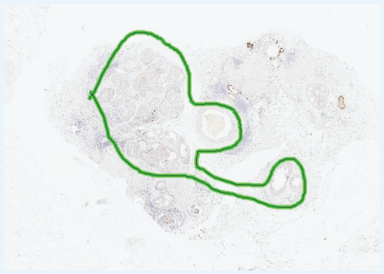


therapies[4,5]. Ki67 is a measure of proliferation and can be used for decisions related to chemotherapy. Classification guidelines vary[6]. Expression of Her2 is given as 0 (no staining observed or incomplete staining of <10% cells), 1+ (incomplete staining of >10% cells), 2+ (incomplete or faint circumferential staining of >10% cells or intense circumferential staining of <10% cells) or 3+ (intense, complete circumferential staining of >10% cells). Score of 0 or 1+ gives a negative classification. A 2+ score is equivocal. Tumors with a 3+ score are positive and are treated with trastuzumab or lapatinib[2].

In the case of an upgrade or change in automated image analysis systems it is reasonable to compare the outputs of the two systems as an initial step in validating the new system. Complete validation should involve comparison to a pathologist's manual interpretations. The present study compares an internally validated system from Aperio ePathology (Leica Biosystems, Wetzlar, Germany) consisting of the ScanScope AT scanner and ImageScope software to a newly acquired system from Ventana Medical Systems (Roche, Tucson, AZ) consisting of the iScan Coreo scanner and Virtuoso software.

METHODS

Tissue sections were from core needle or excisional biopsies performed on twenty patients with invasive breast carcinoma between January 2013 and February 2014. Tissue sections were formalin fixed and paraffin embedded (10% buffered formalin for at least 6 hours and up to a maximum 72 hours for ER, PR, and Ki67 and 48 hours for Her2). Staining was accomplished with the Ventana Benchmark XT following TJUH IHC staining protocols. Sections were imaged with both scanners. A patholigist used ScanScope to select tumor regions appropriate for target quantification. Selection areas were mimicked in Ventana's Virtuoso. Side by side monitors allowed for the mimicked selection areas to be hand drawn with a constant view of the ScanScope selection area.



Pathologist's ImageScope selection

Mimicked Virtuoso selection

[1] Hammond ME et al (*2010*) ASCO/CAP Guideline Recommendations for Immunohistochemical.*Arch of Path & Lab Med* 134: 7:e48-e72. **[2]** Wolff AC et al (2007) ASCO/CAP guideline recommendations for human epidermal growth factor receptor 2. *J of Clin Onc* 25:118–145 **[3]** Francis GD et al (2007) Frequency and reliability of oestrogen receptor. *J of Clin Path* 60:1277–1283

RESULTS

There was high classification agreement given boundary values of ER 1%, PR 1%, Ki67 10%, and Her2 3+.

Her2		Ventana		
		Neg.	Equiv.	Pos.
Aperio	Neg.	12	0	0
	Equiv.	1	3	0
	Pos.	0	0	4

ER		Ventana	
		Neg.	Pos.
Aperio	Neg.	2	0
	Pos.	0	18

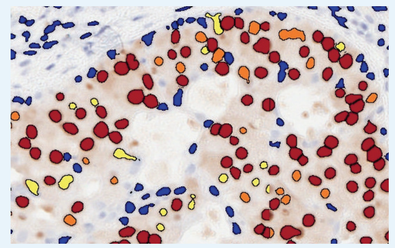
PR		Ventana	
		Neg.	Pos.
Aperio	Neg.	5	1
	Pos.	0	14

Target	Total Agr. (95%CI)	Positive Agr. (95%CI)	Negative Agr. (95%CI)	PPV	NPV
ER	100 (83.2-100)	100 (81.5-100)	100 (15.8-100)	100	100
PR	95.0 (75.1-99.9)	100 (76.8-100)	83.3 (35.9-99.6)	93.3	100
Ki67	90.0 (68.3-98.8)	100 (79.4-100)	50 (6.8-93.2)	88.9	100
Her2	100 (83.2-100)	100 (39.8-100)	100 (79.4-100)	100	100

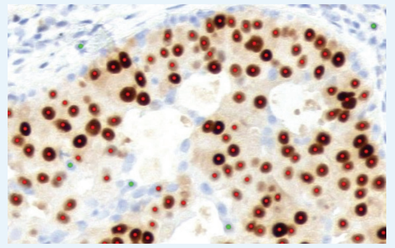
Aperio and Ventana percent positivity scores for ER, PR and Ki67 were different by an average magnitude of 46 (s.d. 49), 67 (65) and 70 (69) percent of the Aperio scores.

CONCLUSIONS

Tumor classification agreement was high suggesting that choice of system has minimal effect on clinical decision making. Large variation in percent positivity scores may indicate suboptimal performance. Visual inspection of software post-analysis images shows that while the Aperio system tends to identify stromal cells as negative tumor cells, the Ventana system undercounts negative tumor cells. An effort is being made to optimize the hematoxylin counterstain for increased negative tumor cell recognition in the Ventana system.



Cell labeling in ImageScope (blue cells are negative)



Cell labeling in Virtuoso (green cells are negative)

[4] EBCTCG (2005) Effects of chemotherapy and hormonal. *Lancet* 365:1687–1717 **[5]** Dowsett M et al (2008) Relationship between quantitative estrogen. *J of Clin Onc* 26:1059–1065 **[6]** Pathmanathan N and Belleine RL (2013) Ki67 and proliferation. *J of Clin Path* 66:512–516