Predicting Phenotypes of Osteoarthritis Progression

Eni Halilaj, Stanford University; *Ya Le, Stanford University; Jennifer L. Hicks, Stanford University; Trevor J. Hastie, Stanford University; Scott L. Delp, Stanford University*

Osteoarthritis (OA) is one of the leading causes of disability in older adults. The uncharacterized heterogeneity of this disease remains a confounder in the design of case-control studies and has significantly hindered the discovery of new disease-modifying treatments. The goal of this study was to characterize distinct phenotypes of osteoarthritis (OA) progression and build a model to predict future progression phenotype based on data collected in one visit. To model structural and symptomatic OA progression, we used eight-year data from the Osteoarthritis Initiative—specifically, joint space width measurements from X-rays and pain scores from the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) questionnaire, since they are the most widely used surrogates of structural and symptomatic disease status. We used a mixed-effects mixture model to cluster the functional data and gradient-boosted decision trees (GBT) to build a cross-validated model for predicting progression phenotypes. The analysis was restricted to 1243 subjects who at the enrollment visit were classified as being at high risk of developing OA based on age, BMI, and medical and occupational histories. The majority of subjects (78%) were slow structural progressors or non-progressors, while the rest (22%) were fast progressors, losing nearly 10% of their joint space width per year. Pain progression was highly variable and not associated with joint space narrowing, thus clustering was based primarily on joint space narrowing. Fast structural progression could be predicted with moderate accuracy based on a GBT model of only a few subject characteristics, including weight, sex, smoking, and history of osteoporosis. Cohort selection based on accurate predictive models will make clinical trials more efficacious, significantly reducing the time and cost needed to detect new anabolic or anti-catabolic drug effects. The discovery of more sensitive imaging and biospecimen markers will also improve the performance of these predictive models in the future.