



# Shrinking the Psoriasis Assessment Gap: Early Gene-Expression Profiling Accurately Predicts Response to Long-Term Treatment

Joel Correa da Rosa<sup>1</sup>, Jaehwan Kim<sup>1,2</sup>, Suyan Tian<sup>3</sup>, Lewis E. Tomalin<sup>2,4</sup>, James G. Krueger<sup>2</sup> and Mayte Suárez-Fariñas<sup>1,2,4,5,6</sup>

There is an “assessment gap” between the moment a patient’s response to treatment is biologically determined and when a response can actually be determined clinically. Patients’ biochemical profiles are a major determinant of clinical outcome for a given treatment. It is therefore feasible that molecular-level patient information could be used to decrease the assessment gap. Thanks to clinically accessible biopsy samples, high-quality molecular data for psoriasis patients are widely available. Psoriasis is therefore an excellent disease for testing the prospect of predicting treatment outcome from molecular data. Our study shows that gene-expression profiles of psoriasis skin lesions, taken in the first 4 weeks of treatment, can be used to accurately predict (>80% area under the receiver operating characteristic curve) the clinical endpoint at 12 weeks. This could decrease the psoriasis assessment gap by 2 months. We present two distinct prediction modes: a universal predictor, aimed at forecasting the efficacy of untested drugs, and specific predictors aimed at forecasting clinical response to treatment with four specific drugs: etanercept, ustekinumab, adalimumab, and methotrexate. We also develop two forms of prediction: one from detailed, platform-specific data and one from platform-independent, pathway-based data. We show that key biomarkers are associated with responses to drugs and doses and thus provide insight into the biology of pathogenesis reversion.

*Journal of Investigative Dermatology* (2017) 137, 305–312; doi:10.1016/j.jid.2016.09.015

## INTRODUCTION

The development of drugs targeting key immune function modulators has revolutionized the treatment of many diseases (Gandhi et al., 2016). A key challenge when assessing the clinical efficacy of drugs targeting specific human immune components is extrapolating from in vitro data and animal models. Although modification of immune components can occur relatively soon after drug administration, clinical responses resulting from these modifications can require months to manifest. This results in an “assessment

gap” between the moment treatment response is determined biologically and when response can actually be assessed clinically. The cost of bringing a successful drug to market is estimated at 2.6 billion dollars and takes around 8 years (Mestre-Ferrandiz et al., 2012; Paul et al., 2010). The prolonged assessment gap associated with immune-targeting drugs necessitates longer trials with human patients, further adding to the cost, complexity, and development time of these drugs. In addition to drug development issues, the prolonged assessment gap impedes primary physicians from selecting the most efficacious treatment from the drug options available. With treatment efficacies varying from 20–80%, multiple treatment attempts are required before the most effective treatment is found for each patient. The prolonged assessment gap can therefore result in patients spending several months using ineffective treatments, increasing costs and adverse effect exposure.

The assessment gap is the “theoretically avoidable” part of the overall time from drug administration to accurate evaluation of treatment response. Upon administration, the time taken for the drug to reach its biological target is the drug delivery time. Once the drug has reached its targets, the biological response time is the time taken for the targets to propagate their influence through the pathogenic pathway until the system is either committed to recovery or, alternatively, the drug has failed to alter pathogenic fate. Collectively, drug delivery and biological response time constitute the assessment gap, during which it could be hypothesized that measurable information exists in the system to enable determination of an individual patient’s response to the treatment. Supporting this hypothesis using an explicit

<sup>1</sup>The Center for Clinical and Translational Science, The Rockefeller University, New York, New York, USA; <sup>2</sup>Laboratory for Investigative Dermatology, The Rockefeller University, New York, NY, USA; <sup>3</sup>Division of Clinical Epidemiology, First Hospital of Jilin University, Changchun, Jilin, China; <sup>4</sup>Department of Population Health Science and Policy, Icahn Institute for Genomics and Multiscale Biology, New York, New York, USA; <sup>5</sup>Department of Genetics and Genomics Science, Icahn Institute for Genomics and Multiscale Biology, New York, New York, USA; and <sup>6</sup>Department of Dermatology, Icahn School of Medicine at Mount Sinai, New York, New York, USA

Correspondence: Mayte Suárez-Fariñas, Department of Population Health Science and Policy and Department of Genetics and Genomics Science, Icahn Institute for Genomics and Multiscale Biology, Department of Dermatology, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, New York, New York 10029. E-mail: [Mayte.SuarezFarinas@mssm.edu](mailto:Mayte.SuarezFarinas@mssm.edu)

Abbreviations: AUC, area under the receiver operating characteristic curve; PASI, Psoriasis Area and Severity Index; PASI75, at least 75% reduction in Psoriasis Area and Severity Index score; UPSC, universal psoriasis skin classifier

Received 21 March 2016; revised 1 September 2016; accepted 3 September 2016; accepted manuscript published online 22 September 2016; corrected proof published online 5 November 2016

demonstration that such measurements and determinations can indeed be performed, with great accuracy, is the subject of this study.

To carry out this demonstration we chose psoriasis, a T-cell-mediated skin disease, as a suitable disease/pathway on which to test this hypothesis. In the last decade, the availability of high-throughput technologies for comprehensive molecular characterization of the disease has greatly increased our understanding of psoriasis. This has resulted in the development of multiple biological therapies that have consistently outperformed earlier treatments. Thus, psoriasis has become a popular immune disease for proof-of-concept studies for targeted immune therapeutics in humans. Furthermore, psoriasis patients display low placebo responses, high potential for therapy-induced disease reversal, and easy accessibility of diseased tissue for associated pharmacodynamics or biomarker analyses. These factors make psoriasis an ideal disease on which to test our hypothesis.

Although psoriasis is highly responsive to pathway-specific immune antagonists, it often takes 12–16 weeks for clinical response measures to be meaningful, and efficacy ranges from 30–80% success rate. Despite the increase in FDA-approved treatments for psoriasis, 20–30% of patients still fail to respond to biologics (Villanova et al., 2013); their effectiveness in real-world practice is lower than that reported in clinical trials (Gelfand et al., 2012). Even specialized dermatologists have difficulties selecting the most effective treatment for individual patients; thus, maximum clinical response is not always achieved (Abuabara et al., 2012; Sequeira et al., 2012). Physicians recommend treatment options on the basis of results of randomized controlled trials and evidence syntheses such as meta-analysis (Timmermans and Mauck, 2005). Although this is currently the best information we can give to patients, there are limitations to this approach. For example, treatment efficacy may not appear significant when averaged across a large cohort of patients, but the treatment may be highly beneficial for a subset of patients. Additionally, not all treatments are compared head to head in randomized controlled trials, so it is difficult to predict which treatment has superior efficacy. The ideal situation would be a test capable of predicting, before commencing therapy, which treatment is most likely to be beneficial for an individual patient. Even if accurate predictions of treatment response cannot be made before treatment, a test capable of predicting disease outcome after a very short treatment would still be a useful tool for decreasing the assessment gap. There is currently tremendous interest in such “stratified medicine” methods, including identification of biomarkers indicative of the most appropriate drug/dose for an individual patient.

In this study, we use machine-learning techniques to analyze the abundant high-throughput molecular data available for psoriasis and generate detailed “molecular phenotypes” that predict drug efficacy with shorter clinical trials and smaller cohorts (see [Supplementary Figure S1](#) online). In this context, we built a classifier combining elements from statistical and machine learning. This classifier uses data, accumulated in the first few weeks of treatment, to accurately predict the expected treatment outcome after several months of drug exposure. Because the prediction is a

categorical outcome (response or no response), we adopt the same terminology used in machine learning areas and will refer to this tool indistinctively as “predictor” or “classifier.”

We set out to develop a classifier to estimate the efficacy profile of new psoriasis drugs from short-term proof-of-concept studies, thus decreasing clinical trial costs. We also envisage uses for rapid assessment of whether a given patient is responding to a particular treatment. This classifier uses RNA expression data from skin biopsy samples obtained in short-term (4 weeks) clinical trials to evaluate the efficacy profile of a psoriasis drug after a full treatment course (>12 weeks). With longitudinal gene expression profiles obtained pretreatment and at intermediate time points during the short-term trial, the tool can predict which individual patients will respond to treatment at the (still unobserved) endpoint at 12 weeks. The classifier computes the classical efficacy endpoint in psoriasis, the drug’s PASI75 efficacy profile: that is, the percentage of patients with at least 75% reduction in the Psoriasis Area & Severity Index (PASI) score after 12 weeks (or later) of treatment.

The classifier is necessarily agnostic to the specific treatment, and we dubbed it the “universal psoriasis skin classifier” (UPSC). Using similar strategies, we also developed treatment-specific classifiers that show specificities in the discriminant genes and pathways determining response to a specific treatment (see [Supplementary Figure S1](#)). A final objective was to establish the shortest trial that can be used while guaranteeing acceptable accuracy. We therefore developed classifiers using data accrued pretreatment (baseline at week [W] 0) up to W1, W2, or W4 posttreatment and evaluated the gain in accuracy resulting from the additional time points.

## RESULTS

To develop and evaluate the classifiers, we used longitudinal skin expression profiles from 141 patients treated with etanercept, ustekinumab, adalimumab, methotrexate, SRT2014, tofacitinib and placebo as described in [Table 1](#). We used recently released data from SRT2014 and tofacitinib to evaluate the “out-of-sample” (generalization) performance of the classifiers. All classifiers were constructed following our general strategy outlined in [Figure 1](#) and the [Materials and Methods](#) section. Treatment-agnostic (UPSC) classifiers were aimed at predicting the response of a cohort to drugs. Additionally, treatment-specific classifiers for four treatments (etanercept, ustekinumab, adalimumab and methotrexate) were developed to predict the response of an individual to a specific drug. UPSC and treatment-specific classifiers were built using combinations of baseline (W0), W1, W2, and W4 gene-expression data (platform-specific), with the objective of determining the shortest time course that will achieve accurate prediction of >W12 clinical outcome. Additionally, gene set variation analysis-pathways involved in drug response were used to build platform-independent classifiers. This was an attempt to account for future changes in technology, and variations in technologies used by different research groups.

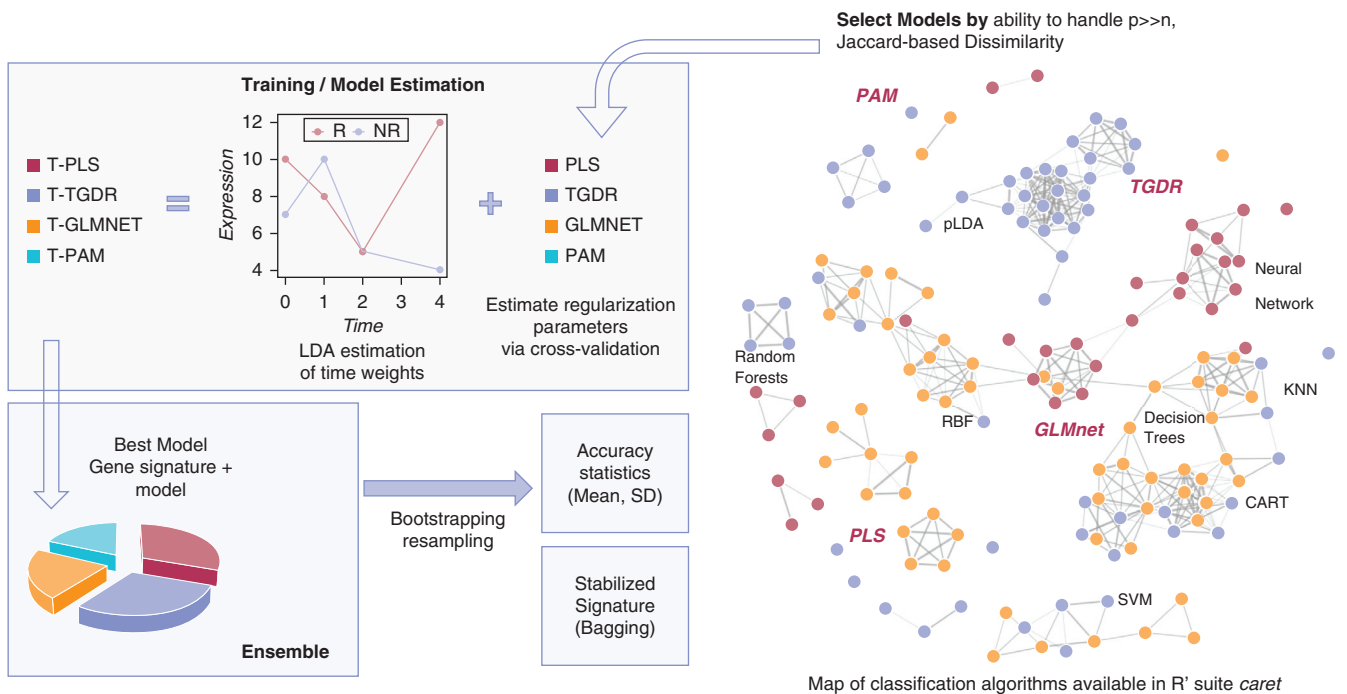
### Classifier performance

**Gene-based classifier.** The data were split into training, for model development (see [Supplementary Table S1](#) online),

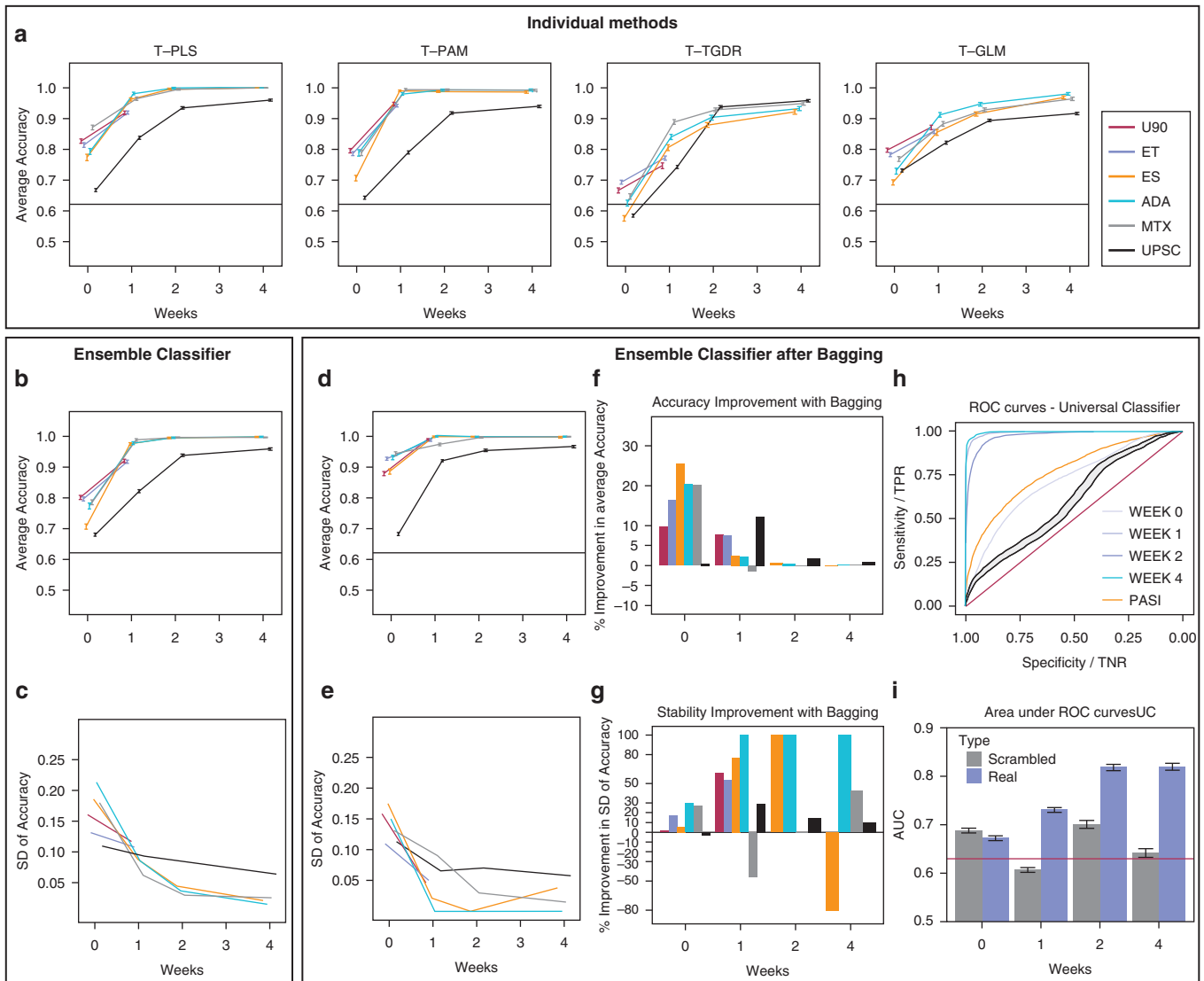
**Table 1. Cohorts used to develop the classifier (training) and evaluate their performance out of sample (generalization)**

| Treatment             | Dosage   | Clinical Trial/Reference<br>GSE Number                               | Nonresponders | Responders | Number<br>of Patients | Time Points<br>(Weeks) | Response endpoint<br>(Week of PASI75<br>Assessment) |
|-----------------------|--|--|---------------|------------|-----------------------|------------------------|---|
| <b>Training</b>       |  |  |               |            |                       |                        |   |
| Etanercept            | 50 mg SC 2× per week for 3 months              | NCT00116181 (Zaba et al., 2009)/GSE11903                             | 4             | 10         | 14                    | 0, 1, 2, 4             | 12  |
| Etanercept            | 50 mg SC 2× per week for 3 months              | NCT01276847 (Suarez-Farinas et al., 2012)/GSE30999                   | 7             | 19         | 26                    | 0, 1                   | 12  |
| Ustekinumab           | 90 mg at W0 and W1                             |  | 8             | 16         | 24                    | 0, 1                   | 12  |
| Adalimumab            | 80 mg SC at W0, 40 mg at W1, and biweekly      | NCT00932113 (Goldminz et al., 2015, 2016)/GSE85034                   | 5             | 10         | 15                    | 0, 1, 2, 4             | 16  |
| Methotrexate          | 7.5 mg per week and increased by protocol      |  | 11            | 4          | 15                    | 0, 1, 2, 4             | 16  |
| <b>Generalization</b> |  |  |               |            |                       |                        |   |
| Ustekinumab           | 45 mg at W0 and W1                             | NCT01276847 (Suarez-Farinas et al., 2012)/W0, GSE30999; W1, GSE85034 | 1             | 15         | 16                    | 0, 1                   | 12  |
| SRT2014               | Placebo, 250 mg, 500 mg or 1000 mg, oral daily | NCT01154101 (Krueger et al., 2015)/GSE50614                          | 2             | 1          | 3                     | 0                      | 12  |
|                       | SRT2104, 250 mg, 500 mg or 1000 mg, oral daily |  | 5             | 2          | 7                     | 0                      | 12  |
| Tofacitinib           | Placebo, 5 mg oral capsules, twice daily       | NCT01710046 (Krueger et al., 2016)/GSE69967                          | 2             | 1          | 3                     | 0, 1, 4                | 12  |
|                       | Tofacitinib, 5 mg oral capsules, twice daily   |  | 2             | 4          | 6                     | 0, 1, 4                | 12  |

Abbreviations: GSE, gene expression data series; SC, subcutaneously; W, week.



**Figure 1. General strategy for building the statistical classifiers.** Dissimilar classification algorithms are selected from those available in the R suite “caret”. We selected PLS, GLMNET, TGDR, and PAM, for their ability to handle  $p \gg n$  and their Jaccard dissimilarity index, and modified them using LDA for application to gene-expression time course data (T-PLS, T-GLMNET, T-TGDR, and T-PAM). The predictions of the final models selected by each algorithm are combined in an ensemble classifier. Performance of homogenous and ensemble classifiers are assessed using a range of methods. Features with highest predictive ability (stabilized gene signature) are selected using a bootstrap aggregate strategy. Models are then built using only the stabilized gene signature to improve accuracy. GLMNET, generalized linear model with convex penalties; LDA, linear discriminant analysis; PAM, prediction analysis for microarrays; PLS, partial least squares; SD, standard deviation; TGDR, threshold gradient descent regularization.



and test, for performance assessment (Figure 2, and see Supplementary Table S1 and Supplementary Figures S2 and S3 online), over 500 iterations. Bootstrap aggregate/bagging was then used to select the most predictive stabilized gene signature, which was used to develop the final models. Classification accuracy of most UPSC classifiers, trained with baseline gene expression, is close to random for most of the methods (Figure 2a, and see Supplementary Figure S2); however, accuracy is increasingly gained with additional time points (Figure 2a, and see Supplementary Figure S2a). This trend is also observed for ensemble classifiers, both pre- (Figure 2b, and see Supplementary Figure S2a) and post-bagging (Figure 2d, and see Supplementary Figure S2b). Furthermore, addition of later time points decreases variation in the accuracy (see the rationale in Supplementary Figure S4

online), producing more consistent results (Figure 2c). Using bagging to select the most predictive genes for the individual classifiers improves the average accuracy of ensemble classifiers, especially for the predictions based on baseline and W1 data (Figure 2d–f). Bagging improves UPSC accuracy 3% at W2 or W4, and halves the size of the signature (see Supplementary Figure S3). Improved stability of prediction is seen after bagging for most classifiers, including a greater than 10% reduction in accuracy standard deviation of the W2 and W4 UPSC (Figure 2g). Stability improvement after bagging was much higher for most treatment-specific classifiers, which also displayed improved accuracy and stability with additional time points (Figure 2a–c).

Overall, these positive results indicate the potential to predict the PASI75 profile of a new compound with greater

than 95% (AUC = 99.9%) on average accuracy after 2 weeks of treatment and greater than 97% (AUC = 99.4%) accuracy after 4 weeks. As for the treatment-specific classifier, the results suggest that we could potentially treat a patient for 1 or 2 weeks and predict the W12 response with more than 97% accuracy (Figure 2d).

To check that these very optimistic results were not due to overfitting, we randomly scrambled our data to produce time course expression data that were essentially noise. We then proceeded to train our classifiers using the same pipeline (see Supplementary Figure S5 online) and repeat the process 500 times. The receiver operating characteristic curves and the AUC (Figure 2h) clearly show the increase in predictive power as more data are accrued and a clearly superior performance compared with the scrambled classifier (whose 95% confidence interval is presented in gray). Much larger AUC values (Figure 2i) are observed for the predictions made by the UPSC than the scrambled classifier ( $P < 10^{-80}$ ), except for baseline predictions, which leads us to reject the overfitting hypothesis.

A predictive model using the clinical measure of disease severity (PASI) from W0–W4, did not have better predictive power than baseline genomic information (Figure 2h), thus reinforcing the idea that patients' molecular profiles can improve prediction of treatment response compared with current clinical measures.

We did consider the possibility that the W4 data were solely responsible for the models' high performances, with the early time points being redundant. However, UPSC classifiers built using either W4 only, or W0 and W4, underperformed by several metrics when compared with those built using the full data set (see Supplementary Table S1 and Supplementary Figure S6 online). Similarly, treatment-specific classifiers built with only W4 data also underperformed compared with those built with additional time points (see Supplementary Table S1 and Supplementary Figure S6).

**Pathway-based classifier.** Pathway-based, platform-independent classifiers were built using the canonical pathways (C2) collection of the MolSigDb (<http://www.broadinstitute.org/gsea/msigdb/collections.jsp#C2>) and a collection of 213 psoriasis-related gene sets curated by our laboratory. This collection includes psoriasis transcriptomes and transcriptional signature induced by IL-17, tumor necrosis factor, IFN- $\gamma$ , and IL-22 on several skin cell types. In general, the pathway-based UPSC did not performed as well as the gene-based UPSC, with accuracy around 0.8 using all data up to W4 (see Supplementary Figure S3a). Notably, bagging does not improve, and in fact decreases, the accuracy of the predictions for the pathway-based UPSC (see Supplementary Figure S3b), and the size of the signature was not greatly reduced (see Supplementary Figure S3c). However, a suitable performance was observed in the treatment-specific classifier, with an average accuracy at W2 of 95% for etanercept, adalimumab and methotrexate and accuracy in testing samples ranging from 79.5–96.0% across all four agents at W1 (see Supplementary Figure S3d).

Most gene set variation analysis-like methods have been developed without consideration of the correlation structure in the data. Although Tarca et al. (2013) showed acceptable

performance for paired data and we chose the methods that better performed in a simulation study (data not shown), a more coherent methodology can be developed, which may potentially lead to improvement in performance.

**Generalization performance.** The ultimate measure of a classifier's success is its ability to make predictions from data that were never used in training. We therefore measured classifier performance against unseen/generalization data for drugs or doses not used during model training (see Supplementary Figure S6 and Supplementary Table S1).

We obtained 25 longitudinal skin expression profiles of various durations for (i) ustekinumab 45 mg ( $n = 16$ ) patients (a lower dosage than the ustekinumab 90 mg used for training the model), (ii) tofacitinib ( $n = 6$ ), an oral Janus kinase inhibitor and (iii) placebo ( $n = 3$ ) from the same trial (Krueger et al., 2016). We also ran predictions using only baseline samples for seven psoriatic patients treated with three doses of SRT2104, a selective activator of SIRT1, and three with placebo (Krueger et al., 2015) (Table 1).

Results for the predictive performance on generalization of the gene- and pathway-based UPSC are presented in Figure 3. Overall, the gene-based classifier predicted treatment response with 84% accuracy, and the pathway-based classifier had 88% accuracy. As expected from our simulation results, there was an underperformance for the predictions for patients treated with SRT2104. The UPSC predicted a PASI75 profile of 33% for placebo, 87.5% for ustekinumab at 45 mg, very close to the real drug profile for that cohort. The optimistic PASI75 prediction for tofacitinib (83.3% prediction vs. 66.7% true response), is an artifact of the small sample size, because the prediction was off by only one patient. The best prediction was produced for the trial with 16 patients, in contrast with the six- to seven-patient cohorts; this may indicate a suitable sample size for this type of study.

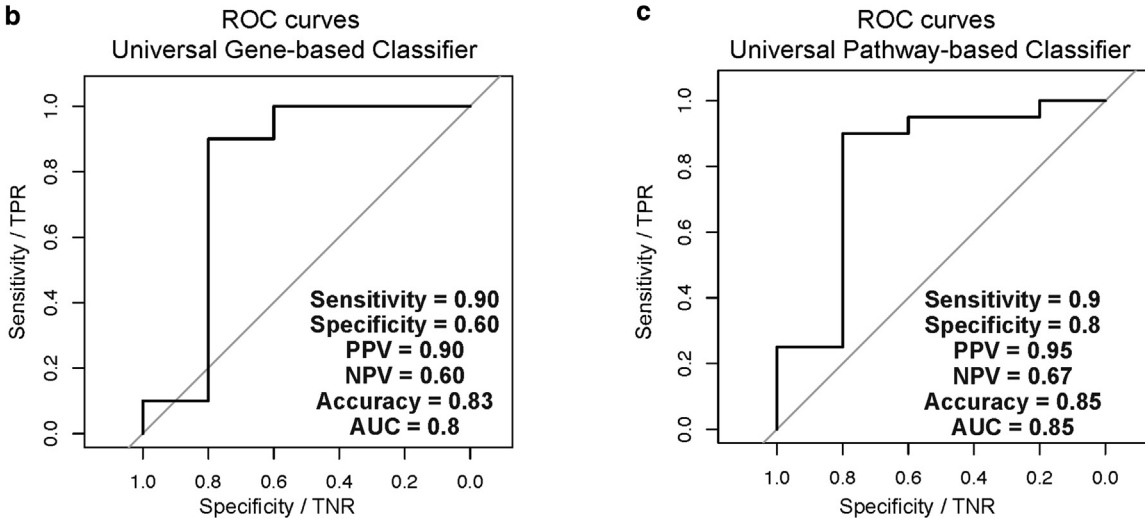
### The biology underlying the UPSC

It is not unusual for machine-learning methods to give accurate classifications or predictions, but when the internal parameters or selected features of the model are examined, no logical interpretation is evident or forthcoming. We nevertheless examined the gene signatures of the classifiers to try to understand the biological basis for their success.

The gene selection for all classifiers was driven by purely statistical/machine-learning criteria without considering the pathogenic role of each gene in psoriasis, to limit selection bias (Barbash and Soreq, 2013). To gain biological insights into the gene-selection mechanism, we performed gene set enrichment analysis with our collection of psoriasis pathways and the canonical C2 pathways. Mirroring the increase in accuracy and stability of UPSC shown in Figure 2, significant enrichment of the predictive genes in psoriasis pathways was achieved as more data points were used (Figure 4 and b). The psoriasis meta-analysis transcriptome/MAD5 (Tian et al., 2012) and genes up-regulated in lesional versus normal skin (Gudjonsson, 2010) were the most significantly enriched pathways (false discovery rate  $< 0.01$ ) among the predictive genes in the ensemble UPSC. Even though gene selection was carried out in an unbiased manner, most of the enriched

**a**

| DP: Drug profile; predicted PASI75 response rate.<br>R: Responders; NR: Non-Responders; A: accuracy of the prediction |                   |                    | Predictions            |                 |                        |                 |
|---|-------------------|--------------------|------------------------|-----------------|------------------------|-----------------|
| Cohort  | Drug              | True Response      | Genes                  |                 | Pathway                |                 |
|   |                   |                    | Prediction/DrugProfile | Accuracy        | Prediction/DrugProfile | Accuracy        |
| MK-0000-206 (NCT01276847)<br>Weeks 0 and 1  | Ustekinumab 45 mg | 15R/1NR (DP=93.8%) | 14R/2NR (DP=87.5%)     | 13/16 (A=81.2%) | 14R/2NR (DP=87.5%)     | 13/16 (A=81.2%) |
| CP-690,550 (NCT01710046)<br>Weeks 0, 1 and 4  | Placebo           | 1R/2NR (DP=33.3%)  | 1R/2NR (DP=33.3%)      | 3/3 (A=100%)    | 1R/2NR (DP=33.3%)      | 3/3 (A=100%)    |
|   | Tofacitinib       | 4R/2NR (DP=66.7%)  | 5R/1NR (DP=83.3%)      | 5/6 (A=83.3%)   | 4R/2NR (DP=66.7%)      | 6/6 (A=100%)    |
| only Baseline values  |                   |                    |                        |                 |                        |                 |
| SRT2104 (NCT01154101)<br>Weeks 0  | Placebo           | 1R/2NR (DP=33.3%)  | 1R/2NR (DP=33.3%)      | 3/3 (A=100%)    | 2R/1NR (DP=66.7%)      | 2/3 (A=66.7%)   |
|   | SRT2104           | 2R/5NR (DP=28.6%)  | 4R/3NR (DP=57.1%)      | 5/7 (A=71.4%)   | 1R/6NR (DP=14.3%)      | 6/7 (A=85.7%)   |



**Figure 3. Performance for the generalization samples.** (a) Summary of results for the generalization samples. (b) Performance statistics for the 25 samples with ROC-curves and (c) performance statistics for the gene-base and pathway-based UPSC for the 25 samples with longitudinal samples. AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating characteristic; TNR, true negative rate.

pathways in Figure 4 are known to be central to disease pathogenesis, including IL-17, tumor necrosis factor, and IFN- $\gamma$  (Lowe et al., 2014).

This trend of identifying pathogenic genes is also highlighted in the final gene signature of the UPSC at W4 (36 genes, see Supplementary Tables S2 and S3), when almost all genes have been identified as involved in psoriasis biology, albeit not the most recognizable “psoriasis genes.” 56% of genes in this signature are part of psoriasis transcriptomes and more than 80% have been shown to display altered expression in response to several psoriasis treatments. Ten genes (28%) are associated with early (2-week) response to a single dose of ixekizumab (Krueger et al., 2012), and three have been directly identified as IL-17 genes in keratinocytes or enhancer of T helper type 17 cells (Vahedi et al., 2015). As mentioned, and despite the striking data mining results, this is not the typical psoriasis biomarker set that experts in the field would put together if requested to do so. It includes as many up-regulated as down-regulated genes, in contrast to the focus on up-regulated genes common in the field.

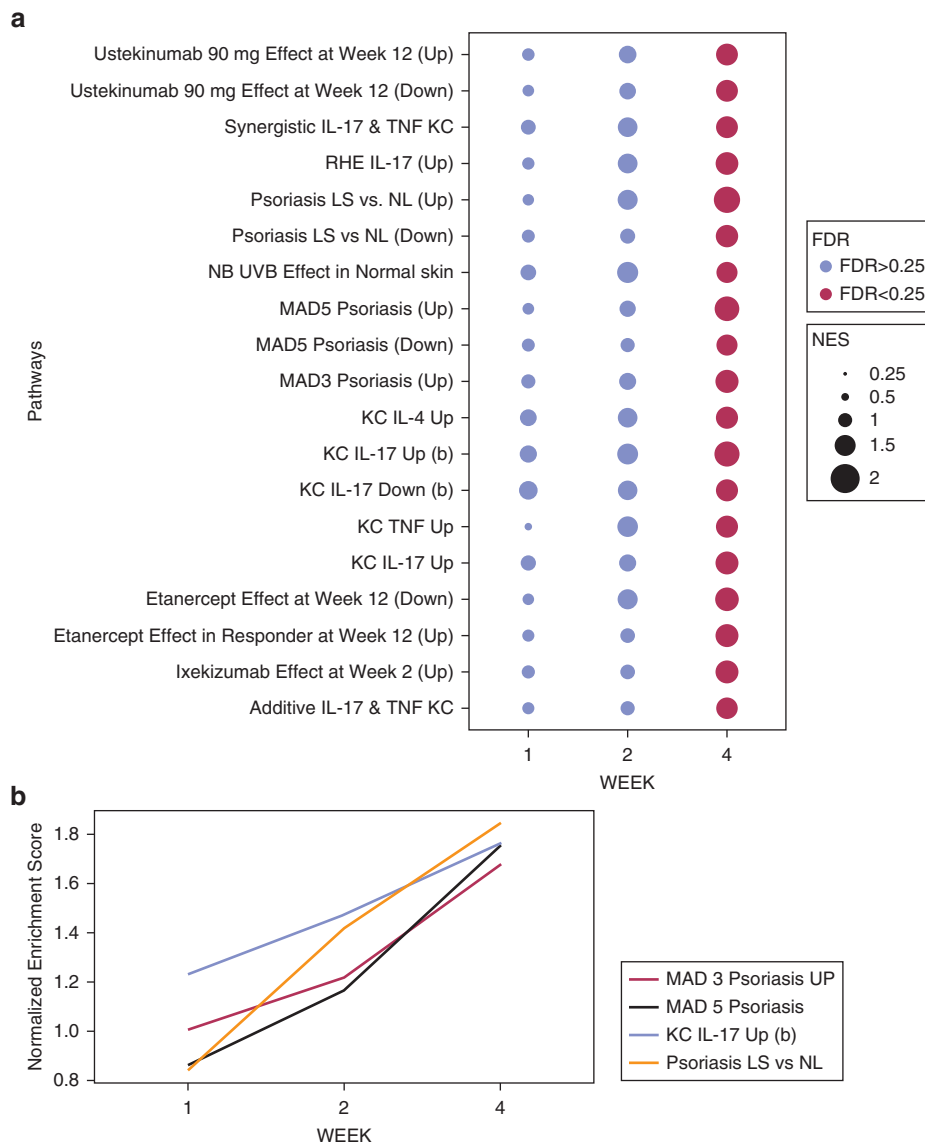
The data support the idea that combining gene expressions from 2–4 weeks posttreatment enhances gene selection to reflect disease pathogenesis more accurately and therefore improves the accuracy and stability of ensemble classifiers.

**DISCUSSION**

As expected on theoretical grounds, the predictive ability of the USPC is poor at baseline, because no information about changes in the molecular phenotype after drug administration is being incorporated into the prediction tool. Despite this fact, at baseline, significantly better than random accuracy was achieved by treatment-specific classifiers, mainly after bagging, showing that patient susceptibility to a specific treatment can be predicted from a gene set and pave the way for the creation of personalized medicine allocation schemes.

A robust and accurate predictor of PASI75 response after 12 weeks of treatment was built using molecular profiling of short-term (2–4 weeks) treatment. We can classify specific treatment response to etanercept, adalimumab and methotrexate with 95% accuracy at W2 and almost 100% accuracy at W4. The accuracy of the UPSC (agent agnostic classifier) was 95% with 2 weeks of data and 97% accuracy with 4 weeks of data. Although ensemble classifiers performed best overall, individual classifiers had similar performance to ensembles. Hallmark psoriasis pathways, including IL-17 became increasingly more relevant at W2 and W4.

Furthermore, these results survived generalization to a new dataset without any retraining. The performance of the UPSC was evaluated in previously unused data from 26 patients treated with ustekinumab 45mg, tofacitinib, or placebo.



**Figure 4. Psoriasis pathways enriched in the genes selected by the ensemble universal psoriasis skin classifier.**

Genes were ranked by the sum of bagging frequencies in each single classifier, and enrichment was evaluated for a curated set of 213 psoriasis pathways. (a) Gene set enrichment analysis normalized enrichment scores obtained after ranking genes by bagging frequency considering the ensemble classifiers. False-discovery rates were used to indicate the significance of association between the ranked gene sets and the pathways. A cutoff of 0.25 was used according to the rationale presented in [Subramanian et al. \(2005\)](#). (b) Normalized enrichment score evolution over time for pathways that showed a false-discovery rate of less than 0.01. FDR, false-discovery rate; KC, keratinocytes; LS, lesional; MAD, meta analysis derived; NES, normalized enrichment scores; NL, nonlesional; RHE, reconstructed human epidermis; TNF, tumor necrosis factor.

Overall, 83% accuracy and AUC of 0.83 were obtained for the gene-based classifier and 85% accuracy and AUC of 0.85 for the pathway-based classifier, with higher accuracies for the nine patients with complete longitudinal profiles until W4 (88.8% and 100% for the gene- and pathway-based classifiers, respectively).

This study has demonstrated that the assessment gap can indeed be considerably shortened in psoriasis by showing that molecular data from the first 4 weeks can be used to accurately predict the clinical outcome at W12. The tool we used for this demonstration is a set of predictors built through rigorous guiding principles of robustness and platform independence. Our implemented pipeline was very conservative, because we aimed to minimize biases that could overestimate performance. We have shown that they can be generalized to different drugs or doses without retraining and that the features they extracted for prediction accurately recap the pathogenesis of the disease. Although algorithms based on smaller signatures may lead to classifiers with adequate performance ([Saez-Rodriguez et al., 2016](#)), it is reassuring that despite its complexity, our unbiased feature

selection generally leads to signatures with fewer than 40 genes (see [Supplementary Tables S2 and S3](#)).

This demonstration gives us hope that our methodological and computational framework can be used for other chronic inflammatory diseases and for personalized medicine use. Our analysis shows that the biology of pathogenesis recovery is accurately summarized and suggests that using this approach with diseases of ill-understood pathogenesis may help highlight genes of direct biological significance to its pathogenic pathways, aiding basic science and suggesting possible new interventional targets.

The highly promising and encouraging results we obtained with our prediction tool in psoriasis emphasize the potential of molecular profiling coupled with machine learning and statistical techniques to accelerate and optimize the conduction of clinical trials. Finally, because psoriasis is arguably the model immune-mediated skin disease, we hope that the lessons learned here in terms of closing the assessment gap can be extended to other common inflammatory skin diseases such as atopic dermatitis, alopecia areata, and lichen planus, that affect millions of people worldwide.

## MATERIALS AND METHODS

### Data

The expression data used in this study were either from previously published research (Krueger et al., 2010) or shared by collaborators; they included phenotypic information and microarray gene expression data (obtained using Affymetrix and Illumina platforms) of lesional and nonlesional skin of psoriatic patients taken at baseline and after 1, 2, and 4 weeks of treatment (Table 1). The response to treatment was measured by the PASI75 endpoint, at which patients are considered responders if there was an improvement of at least 75% on the PASI score from baseline and are considered nonresponders otherwise.

### General framework for classifier construction

Our general strategy for building the statistical classifiers is outlined in Figure 1. Data from patients treated with etanercept, ustekinumab, adalimumab, and methotrexate were split into training (for model development) and test (for evaluating classifier performance). Heterogeneous classification algorithms used for training were selected from the R suite “caret” for their ability to handle  $p \gg n$  and high Jaccard dissimilarity index (PLS, partial least squares; GLMNET, generalized linear models with elastic net; TGDR, threshold gradient descent regularization; and PAM, prediction analysis of microarrays). These algorithms were modified using linear discriminant analysis to allow their application to gene-expression time-course data. The predictions of the final models selected by each algorithm are combined in an ensemble classifier. Bagging was used to select the most predictive stabilized gene signature on the basis of frequency of selection in each bootstrapping iteration model. Final individual and ensemble models were built using this gene signature, the performance of which was assessed against training and test data, as well as the out-of-sample data from SRT2014 and tofacitinib.

A more detailed description of classifier construction, classification algorithms, methodology for handling time course microarray gene expression, model estimation, feature selection, and performance evaluation is given in the [Supplementary Materials and Methods](#).

### CONFLICT OF INTEREST

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this study. JGK has been a consultant to and receives research support from Amgen, Boehringer, Centocor/Janssen, Merck, Pfizer, Idera, and Astellas. MSF receives research support from Quorum Consulting and Pfizer.

### ACKNOWLEDGMENTS

We thank Marcelo O. Magnasco and Gustavo Stolovitzky for their insightful discussions. JCR, JK, and MSF are supported by the Rockefeller University Center for Clinical and Translational Science grant no. UL1 TR000043 from the National Center for Advancing Translational Sciences, National Institutes of Health Clinical and Translational Science Award program. MSF is supported by Irma T. Hirsch/Monique Weill-Caulier Career Scientist Award.

### AUTHOR CONTRIBUTIONS

JCR designed the study, established and performed analysis, and cowrote the article. JK and LT contributed to analysis and writing. ST contributed to data analysis and tools. JGK supervised and codesigned the study. MSF conceptualized and designed the study, directed the study, cowrote the article.

### SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at [www.jidonline.org](http://www.jidonline.org), and at <http://dx.doi.org/10.1016/j.jid.2016.09.015>.

### REFERENCES

Abuabara K, Freeman EE, Dellavalle R. The role of systematic reviews and meta-analysis in dermatology. *J Invest Dermatol* 2012;132:e2.  
Barbash S, Soreq H. Statistically invalid classification of high throughput gene expression data. *Sci Rep* 2013;3:1102.

Gandhi NA, Bennett BL, Graham NM, Pirozzi G, Stahl N, Yancopoulos GD. Targeting key proximal drivers of type 2 inflammation in disease. *Nat Rev Drug Discov* 2016;15:35–50.  
Gelfand JM, Wan J, Duffin KC, Krueger GG, Kalb RE, Weisman JD, et al. Comparative effectiveness of commonly used systemic treatments of phototherapy for moderate to severe plaque psoriasis in the clinical practical setting. *Arch Dermatol* 2012;148:487–94.  
Goldminz AM, Suarez-Farinas M, Wang AC, Dumont N, Krueger JG, Gottlieb AB. CCL20 and IL22 messenger RNA expression after adalimumab vs methotrexate treatment of psoriasis: a randomized clinical trial. *JAMA Dermatol* 2015;151:837–46.  
Goldminz AM, Suarez-Farinas M, Wang AC, Dumont N, Krueger JG, Gottlieb AB. Methotrexate improves pro- and anti-atherogenic genomic expression in psoriatic skin. *J Dermatol Sci* 2016;82:207–9.  
Gudjonsson ED, Johnston A, Tejasvi T, Guzman AM, Nair RP, Vorhees JJ, Abecasis G, Elder JT. Assessment of the psoriatic transcriptome in a large sample: additional regulated genes and comparisons with in vitro models. *J Invest Dermatol* 2010;130:1829–40.  
Krueger J, Brodmerkel C, Li K, Suarez-Farinas M. The molecular profile of psoriatic skin in responders to ustekinumab or etanercept after 12 weeks of treatment: results from the ACCEPT trial. *J Am Acad Dermatol* 2010;62 (3 Suppl. 1):AB13.  
Krueger J, Clark JD, Suarez-Farinas M, Fuentes-Duculan J, Cueto I, Wang CQ, et al. Tofacitinib attenuates pathologic immune pathways in patients with psoriasis: a randomized phase 2 study. *J Allergy Clin Immunol* 2016;137:1079–90.  
Krueger JG, Fretzin S, Suárez-Fariñas M, Haslett PA, Phipps KM, Cameron GS, et al. IL-17A is essential for cell activation and inflammatory gene circuits in psoriasis. *J Allergy Clin Immunol* 2012;130:145–54.  
Krueger JG, Suárez-Fariñas M, Cueto I, Khacherian A, Matheson R, Parish LC, et al. A randomized, placebo-controlled study of SRT2104, a SIRT1 activator, in patients with moderate to severe psoriasis. *PLoS One* 2015;10:e0142081.  
Lowes MA, Suárez-Fariñas M, Krueger JG. Immunology of psoriasis. *Ann Rev Immunol* 2014;32:227–55.  
Mestre-Ferrandiz J, Sussex J, Towse A. The R&D cost of a new medicine. London, UK: Office of Health Economics; 2012.  
Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* 2010;9:203–14.  
Saez-Rodriguez J, Costello JC, Friend SH, Kellen MR, Mangravite L, Meyer P, et al. Crowdsourcing biomedical research: leveraging communities as innovation engines. *Nat Rev Genet* 2016;17:470–86.  
Sequeira VB, Rybchyn MS, Tongkao-on W, Gordon-Thomson C, Malloy PJ, Nemere I, et al. The role of the vitamin D receptor and ERp57 in photoprotection by 1 $\alpha$ , 25-dihydroxyvitamin D3. *Mol Endocrinol* 2012;26:574–82.  
Suarez-Farinas M, Li K, Fuentes-Duculan J, Hayden K, Brodmerkel C, Krueger JG. Expanding the psoriasis disease profile: interrogation of the skin and serum of patients with moderate-to-severe psoriasis. *J Invest Dermatol* 2012;132:2552–64.  
Subramaniana A, Tamayoa P, Moothaa VK, Mukherjeed S, Eberta BL, Gillettea MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005;102:15545–50.  
Tarca AL, Bhatti G, Romero R. A comparison of gene set analysis methods in terms of sensitivity, prioritization and specificity. *PLoS One* 2013;8:e79217.  
Tian S, Krueger JG, Li K, Jabbari A, Brodmerkel C, Lowes MA, et al. Meta-analysis derived (MAD) transcriptome of psoriasis defines the “core” pathogenesis of disease. *PLoS One* 2012;7:e44274.  
Timmermans S, Mauck A. The promises and pitfalls of evidence-based medicine: nonadherence to practice guidelines remains the major barrier to the successful practice of evidence-based medicine. *Health Affairs* 2005;24:18–28.  
Vahedi G, Kanno Y, Furumoto Y, Jiang K, Parker SCJ, Erdos MR, et al. Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature* 2015;520:558–62.  
Villanova F, Meglio PD, Nestle FO. Biomarkers in psoriasis and psoriatic arthritis. *Ann Rheum Dis* 2013;72(Suppl. 2):104–10.  
Zaba LC, Suarez-Farinas M, Fuentes-Duculan J, Nograles KE, Guttman-Yassky E, Cardinale I, et al. Effective treatment of psoriasis with etanercept is linked to suppression of IL-17 signaling, not immediate response TNF genes. *J Allergy Clin Immunol* 2009;124:1022–10.e1-395.