Using molecular classification to predict gains in maximal aerobic capacity following endurance exercise training in humans

James A. Timmons,[1,2,3] **Steen Knudsen,**[4] **Tuomo Rankinen,**[5] **Lauren G. Koch,**[6] **Mark Sarzynski,**[5] **Thomas Jensen,**[4] **Pernille Keller,**[7,8] **Camilla Scheele,**[3,8] **Niels B. J. Vollaard,**[9] **Søren Nielsen,**[8] **Thorbjörn Åkerström,**[8] **Ormond A. MacDougald,**[7] **Eva Jansson,**[10] **Paul L. Greenhaff,**[11] **Mark A. Tarnopolsky,**[12] **Luc J. C. van Loon,**[9] **Bente K. Pedersen,**[8] **Carl Johan Sundberg,**[13] **Claes Wahlestedt,**[14] **Steven L. Britton,**[6,7] **and Claude Bouchard**[5]

[1]*Section of Systems Biology Research, Panum Institutet and Center for Healthy Ageing,* [8]*Centre of Inflammation and Metabolism, Faculty of Health Sciences, University of Copenhagen, Copenhagen,* [4]*Medical Prognosis Institute, Hørsholm, Denmark;* [2]*Royal Veterinary College, University of London, Camden, London,* [11]*Centre for Integrated Systems Biology Medicine, University Medical School, Nottingham, United Kingdom;* [3]*The Wenner-Gren Institute, Arrhenius Laboratories, Stockholm University,* [10]*Department of Laboratory Medicine, Division of Clinical Physiology, Karolinska University Hospital,* [13]*Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden;* [5]*Human Genomics Laboratory, Pennington Biomedical Research Center, Baton Rouge, Louisiana; Departments of* [6]*Physical Medicine and Rehabilitation and* [7]*Molecular and Integrative Physiology, University of Michigan Medical School, Ann Arbor, Michigan;* [9]*Department of Human Movement Sciences, Nutrition and Toxicology Research Institute Maastricht, Maastricht University Medical Centre, Maastricht, The Netherlands;* [12]*Department of Paediatrics and Medicine (Neurology and Rehabilitation), McMaster University Medical Centre, Hamilton, Ontario, Canada; and* [14]*Molecular and Integrative Neurosciences Department, The Scripps Research Institute, Jupiter, Florida*

Timmons JA, Knudsen S, Rankinen T, Koch LG, Sarzynski M, Jensen T, Keller P, Scheele C, Vollaard NB, Nielsen S, Åkerström T, MacDougald OA, Jansson E, Greenhaff PL, Tarnopolsky MA, van Loon LJ, Pedersen BK, Sundberg CJ, Wahlestedt C, Britton SL, Bouchard C. Using molecular classification to predict gains in maximal aerobic capacity following endurance exercise training in humans. *J Appl Physiol* 108: 1487–1496, 2010. First published February 4, 2010; doi:10.1152/japplphysiol.01295.2009.—A low maximal oxygen consumption ($\dot{V}O_{2max}$) is a strong risk factor for premature mortality. Supervised endurance exercise training increases $\dot{V}O_{2max}$ with a very wide range of effectiveness in humans. Discovering the DNA variants that contribute to this heterogeneity typically requires substantial sample sizes. In the present study, we first use RNA expression profiling to produce a molecular classifier that predicts $\dot{V}O_{2max}$ training response. We then hypothesized that the classifier genes would harbor DNA variants that contributed to the heterogeneous $\dot{V}O_{2max}$ response. Two independent preintervention RNA expression data sets were generated ($n = 41$ gene chips) from subjects that underwent supervised endurance training: one identified and the second blindly validated an RNA expression signature that predicted change in $\dot{V}O_{2max}$ ("predictor" genes). The HERITAGE Family Study ($n = 473$) was used for genotyping. We discovered a 29-RNA signature that predicted $\dot{V}O_{2max}$ training response on a continuous scale; these genes contained ~6 new single-nucleotide polymorphisms associated with gains in $\dot{V}O_{2max}$ in the HERITAGE Family Study. Three of four novel candidate genes from the HERITAGE Family Study were confirmed as RNA predictor genes (i.e., "reciprocal" RNA validation of a quantitative trait locus genotype), enhancing the performance of the 29-RNA-based predictor. Notably, RNA abundance for the predictor genes was unchanged by exercise training, supporting the idea that expression was preset by genetic variation. Regression analysis yielded a model where 11 single-nucleotide polymorphisms explained 23% of the variance in gains in $\dot{V}O_{2max}$, corresponding to ~50% of the estimated genetic variance for $\dot{V}O_{2max}$.

In conclusion, combining RNA profiling with single-gene DNA marker association analysis yields a strongly validated molecular predictor with meaningful explanatory power. $\dot{V}O_{2max}$ responses to endurance training can be predicted by measuring a ~30-gene RNA expression signature in muscle prior to training. The general approach taken could accelerate the discovery of genetic biomarkers, sufficiently discrete for diagnostic purposes, for a range of physiological and pharmacological phenotypes in humans.

endurance training; genotype; personalized medicine

LOW EXERCISE AEROBIC CAPACITY is associated with an increased risk of metabolic and cardiovascular disease, as well as premature death. Exercise capacity, in prospective follow-up analyses, is a stronger predictor of morbidity and mortality than other established risk factors, such as hypertension or diabetes (5, 6, 19, 29, 37). A notable observation in the search for relevant mechanisms that connect aerobic capacity with disease is that humans can increase peak oxidative power through regular exercise with a degree of heterogeneity ranging from 0% to >100%, and thus some subjects are unable to improve their aerobic capacity (7, 34, 54). Maximal aerobic capacity is commonly thought to be limited by maximal delivery of oxygen to the periphery and, hence, by cardiac function (44). This being the case, it is somewhat counterintuitive that baseline aerobic capacity neither positively nor negatively associates with the gains in exercise training-induced maximal aerobic power. The estimated heritability for gains in aerobic capacity in response to endurance training is ~50% (10, 11). Discovery of the genetic contribution to the heterogeneous response to training (20, 50) may provide the greatest opportunity to contribute to a theoretical molecular framework to link determinants of oxidative capacity to disease pathways and to discover the molecular determinants of this critical physiological system.

We presume that at least part of the heterogeneity in adaptation to exercise therapy originates from variation in gene sequence that influences the complex biological networks mediating the response to an aerobic exercise-training stimulus. Identification of genomic markers for complex traits in humans has required enormous sample sizes, and each single-nucleotide polymorphism (SNP) seems to contribute only weakly, at least for chronic complex human diseases (16). For example, after genome-wide association analysis in type 2 diabetes patients, 18 robust SNPs explain <7% of the total disease variance (46). Enrichment of disease-associated gene networks, generated from RNA expression data, with genotyping data has successfully enhanced the explanatory power of the analysis (13). However, the opposite strategy, where an RNA expression-based molecular classifier (28) is used to locate or rank a discrete number of genes for subsequent targeted genotyping and, hence, discovery of key genetic variants, has not been attempted to our knowledge. We believe that generation of an independently validated RNA-based classifier and then identification of genetic variants in those classifier genes in an independent study not only further validates the molecular classifier, but also exemplifies a strategy to enhance the statistical power of genetic biomarker studies. Thus we hypothesized that baseline RNA expression profiling would identify genes with relevant genetic variants that relate to the large intersubject trait variation observed in humans. Here we apply this rationale to study maximal aerobic power and identify and blindly validate a novel gene signature that predicts the magnitude of adaptation to supervised exercise-training programs.

METHODS

Three independent clinical studies contributed to this project. The first was used to generate an RNA-based molecular classifier, the second to independently validate the RNA classifier, and the third to provide the link between the classifier and genetic variation (Fig. 1). Each clinical study represents a variation on a typical supervised endurance-training program, with subjects of different levels of physical fitness helping establish that our results can be applied broadly to various types of aerobic exercise therapy and subjects.

Group 1. Twenty-four young sedentary healthy Caucasian men (see supplemental Table S1 in the online version of this article) volunteered to take part in the study. Body mass did not change during the study period (78.6 ± 2.7 vs. 78.8 ± 2.6 kg). Mean resting blood pressure (systolic/diastolic) and heart rate were 126/72 mmHg and 66 ± 3 beats/min, respectively. Subjects did not undertake any regular sporting activities in the 6 mo prior to the study and abstained from any strenuous exercise during the 3 wk before pretraining muscle biopsies were obtained. Each of the volunteers provided written informed consent. The study was approved by the Ethics Committee at the Karolinska Institutet. Subjects performed a fully supervised 6-wk training program consisting of four 45-min cycling sessions per week at an intensity customized to 70% of pretraining maximal oxygen consumption ($\dot{V}_{O_{2max}}$; 100% compliance). Physiological measurements and vastus lateralis muscle biopsies were taken prior to the training program and 24 h after the last training session, as previously described (49, 50). $\dot{V}_{O_{2max}}$ was determined during an exhaustive incremental cycling test (Rodby), with continuous analysis of respiratory gases using a SensorMedics ventilator. At $\dot{V}_{O_{2max}}$, the respiratory exchange ratio exceeded 1.10 on all occasions. The clinical study was approved by the Ethics Committee at the Karolinska Institutet.

Group 2. Seventeen young active Caucasian subjects (see supplemental Table S2) volunteered to train on a cycle ergometer (model 839E, Monark, Varberg, Sweden) five times a week for 12 wk. The training load was incrementally customized during the study, such that these active/trained subjects trained at a higher intensity and volume than *group 1* subjects. As part of the training, subjects performed a peak power (P_{max}) test every Monday to determine the intensity of the training for the following days. The P_{max} test was performed the same way as the $\dot{V}_{O_{2max}}$ test without measurement of oxygen consumption ($\dot{V}_{O_2}$). On Tuesdays, the training consisted of ten 3-min intervals at 85% P_{max} separated by 3-min intervals at 40% P_{max}. On the next day, the training consisted of 60 min at 60% P_{max}. On Thursdays, subjects performed five 8-min intervals at 75% P_{max} separated by 4-min intervals at 40% P_{max}. On Fridays, subjects cycled for 120 min at 55% P_{max} continuously. During the first 6 wk, the duration of each training session was increased by 5% each week; during the last 6 wk, the duration remained stable, and the relative intensity was increased 1% per week. This protocol represents the upper limit of intervention one could expect to apply to a general population of healthy and fit subjects. Since our aim was to find genes that contribute universally to aerobic adaptation, we believed that genes deemed regulatory during constant-load (*group 1*) and interval-based (*group 2*) aerobic training would be the best candidates for genotyping. The compliance to training was ~100%. The study was approved by the Ethical Committee of Copenhagen and Frederiksberg Communities, Denmark, and performed according to the Declaration of Helsinki.

HERITAGE Family Study aerobic-training program. The study cohort in this analysis consists of 473 Caucasian subjects (230 male and 243 female) from 99 nuclear families who completed ≥58 of the prescribed 60 exercise-training sessions. The study design and inclusion criteria have been described previously (9) and were established before *group 1* or *2* training formats. To be eligible, the individuals were required to be in good health, i.e., free of diabetes, cardiovascular diseases, or other chronic diseases that would prevent their participation in an exercise-training program. Subjects were also required to be sedentary, defined as not having engaged in regular physical activity over the previous 6 mo. Individuals with resting systolic blood pressure >159 mmHg and/or diastolic blood pressure >99 mmHg or taking medication for hypertension, dyslipoproteinemia, or hyperglycemia were excluded. Other exclusion criteria are fully described in a previous publication (9). The prevalence of overweight and obesity was 30.8% and 19.3%, respectively. The study protocol had been approved by each of the Institutional Review Boards of the HERITAGE Family Study research consortium. Written informed consent was obtained from each participant.

The exercise intensity of the 20-wk program was customized for each participant based on the heart rate (HR)-$\dot{V}_{O_2}$ relationship measured at baseline (11). During the first 2 wk, the subjects exercised at a HR corresponding to 55% of the baseline $\dot{V}_{O_{2max}}$ for 30 min per session. Duration and intensity of the sessions were gradually increased to 50 min and 75% of the HR associated with baseline $\dot{V}_{O_{2max}}$, which were then sustained for the last 6 wk. Frequency of
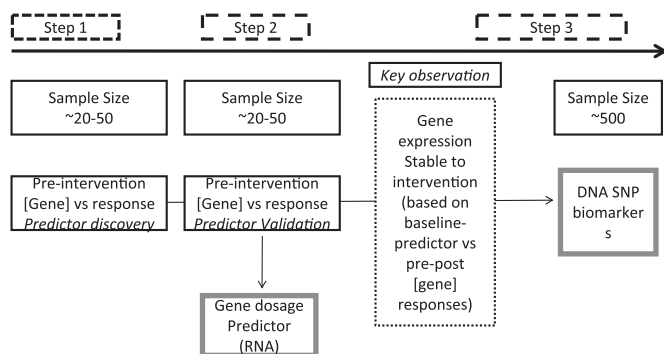

Fig. 1. An overview of the basic research strategy, along with the approximate sample sizes, used in this study. SNP, single-nucleotide polymorphism.

sessions was three times per week, and all exercise was performed on cycle ergometers in the laboratory. HR was monitored during all training sessions by a computerized cycle ergometer system (Universal FitNet System) that adjusted ergometer resistance to maintain the target HR. Trained exercise specialists supervised all exercise sessions. Before and after the 20-wk training program, each subject completed three cycle ergometer (Ergo-Metrics 800S, SensorMedics, Yorba Linda, CA) exercise tests conducted on separate days: a maximal exercise test (Max), a submaximal exercise test (Submax), and a submaximal/maximal exercise test (Submax/Max). The Max test started at 50 W for 3 min, and the power output was increased by 25 W every 2 min thereafter to the point of exhaustion. For older, smaller, or less fit subjects, the test was started at 40 W and increased by 10- to 20-W increments. On the basis of the results of the Max test, the Submax test was performed at 50 W and 60% of the initial $\dot{V}_{O_{2max}}$. Finally, the Submax/Max test was started with the Submax protocol and progressed to a maximal level of exertion. For all tests, $\dot{V}_{O_2}$, CO_2 production, expiratory minute ventilation, and tidal volume were determined every 20 s and reported as a rolling average of the three most recent 20-s values. All the respiratory phenotypes were measured using a SensorMedics 2900 metabolic measurement cart. $\dot{V}_{O_{2max}}$ was defined as the mean of the highest $\dot{V}_{O_2}$ values determined on each of the maximal tests or the higher of the two values if they differed by >5%.

Affymetrix microarray process. Total RNA was extracted from frozen muscle using TRIzol reagent. Frozen pieces were homogenized for 60 s in 1 ml of TRIzol using a 7-mm Polytron aggregate (model PT-DA 2107, Kinematica) adapted to a Polytron homogenizer (model PT-2100) running at maximum speed. RNA concentration and quality were controlled using a NanoDrop spectrophotometer and a bioanalyzer. In vitro transcription (IVT) was made using the Bioarray high-yield RNA transcript labeling kit (P/N 900182, Affymetrix). Unincorporated nucleotides from the IVT reaction were removed using the RNeasy column (Qiagen). *Group 2* IVT was performed using the MessageAmp II biotin-enhanced aRNA kit (Ambion). The effect of the IVT kit was assessed by processing two samples with the Affymetrix kit used for *group 1*. Hybridization, washing, staining, and scanning of the arrays were performed according to the manufacturer's instructions (Affymetrix, http://www.affymetrix.com/). As a means to control the quality of the individual arrays, all were examined using hierarchical clustering and quality control calculations from Bioconductor (e.g., NUSE) to identify outliers prior to statistical analysis, in addition to the standard quality assessments, including scaling factors and housekeeper 5′-to-3′ ratios. The raw data for the predictor data set have been submitted to the Gene Expression Omnibus database (GSE18583). The remaining files will be deposited at GEO following a second article, where we examine the training-responsive muscle mRNA transcriptome (TRT) in detail and contrast the responses across high- and low-responder subjects.

General array analysis methods. The microarray data (GSE18583) were subjected to global normalization using MAS5.0, and present-absent calls were used to improve the sensitivity of the differential gene expression analysis by improving the power while potentially removing some genuinely expressed genes (14). We chose to retain probe sets with ≥25% of the chips declaring a "present" detection, on the basis that there will be subject-to-subject variability and that some genes may be only expressed before or following training. The normalized log2 file was analyzed with the Significance Analysis of Microarray (SAM) in R (http://www-stat.stanford.edu/~tibs/SAM/) (20). SAM provides an estimate of the false discovery rate (FDR), which represents the percentage of genes that could be identified by chance, and is comparable to a *P* value corrected for the number of initial comparisons, a process called multiple testing correction. For the data presented here, genes were considered significantly changed following training, when a delta value corresponding to the number of false significant genes of 5% (*q* value) and an average fold change of 1.5 were achieved. We previously demonstrated that it can be difficult

to predict the impact of applying arbitrary filtering criteria prior to statistical analysis (30). We therefore also relied on several statistical models to analyze and interpret our data. For example, we used the web-based bioinformatics tool INGENUITY pathway analysis (http://www.ingenuity.com), which is based on >1.7 million published PubMed articles and the web-based tool DAVID (Database for Annotation, Visualization, and Integrated Discovery) (15, 25).

Using baseline array data to produce a quantitative predictor of response to training. A quantitative predictor of response to training was developed by correlating measured training-induced change in $\dot{V}_{O_{2max}}$ to RNA expression level of genes in a muscle biopsy obtained prior to training. CEL files were logit-normalized, and an RNA expression index was calculated using model-based analysis of oligonucleotide arrays (32). The probe weights, determined for the training set files, were reused for the test set to increase comparability. To calculate the correlation between $\dot{V}_{O_{2max}}$ response and gene RNA expression for a given gene or probe set, the Pearson correlation for each perfect-match probe in the probe set was calculated, and only the median correlation was retained. If the median correlation of the selected "probe" exceeded 0.3, the entire probe set was retained as originally correlated. We have found that this two-step process helps compensate for the multiple comparisons being made. Correlated probe sets were identified 24 times on the 24-sample training set, each time with 1 sample left out of the calculation. Probe sets were ranked according to the number of times out of 24 (the number of subjects) they were selected as having a median correlation >0.3 (the median value reflecting the probe-level data). The top 29 genes, selected ≥22 of 24 times, gave the best correlation to $\dot{V}_{O_{2max}}$ on the training set (*group 1*). The sum of the RNA expression indexes was calculated and was evaluated using the independent test gene chip data sets derived from *group 2* samples. The logit-normalized model-based RNA expression index (33) values for each of the 29 genes were centered (around 1), and the index was scaled, using all genes, over the 24 subjects in *group 1*. The RNA expression value of the 29 genes was then determined in *group 2*, and the sum of the expression of the 29 genes was correlated to the measured change in $\dot{V}_{O_{2max}}$ by an observer blinded to sample identity. Genes were clustered and plotted using the Heatplus package of Bioconductor (www.bioconductor.org). To allow comparison between cohorts with a different baseline $\dot{V}_{O_{2max}}$, we used the percent change in $\dot{V}_{O_{2max}}$. Finally, for genes and SNPs identified in the HERITAGE Family Study (see below), we validated the genetic association data using RNA expression-based correlation analysis in the (*group 2*) blind validation data set.

Genotype validation of the RNA expression-based predictor. Linkage disequilibrium (LD) cluster-tagging SNPs (tagSNPs) were selected from the Caucasian data set of the International HapMap consortium (data release 23, March 2008). Target areas for the SNP selection were defined as the coding region of the gene plus 20 kb upstream of the 5′ end and 10 kb downstream of the 3′ end of the gene. TagSNPs were selected using the pairwise algorithm of the Tagger program (45). Minor allele frequency was required to be >10%, and the pairwise LD threshold for the LD clusters was set to $r^2 \geq 0.80$.

Genomic DNA was prepared from permanent lymphoblastoid cells by a commercial DNA extraction kit (Gentra Systems, Minneapolis, MN). The tagSNPs were genotyped using Illumina (San Diego, CA) GoldenGate chemistry and Sentrix Array Matrix technology on the BeadStation 500GX. Genotype calling was done with Illumina Bead-Studio software, and each call was confirmed manually. For quality control purposes, each 96-sample array matrix included 1 sample in duplicate, and 47 samples were genotyped in duplicate on different arrays. In addition, six Centre d'Etude du Polymorphisme Humain control DNA samples (NA10851, NA10854, NA10857, NA10859, NA10860, and NA10861; all samples included in the HapMap Caucasian panel) were genotyped. Concordance between the replicates, as well as with the SNP genotypes from the HapMap database, was 100%.

A χ^2 test was used to verify whether the observed genotype frequencies were in Hardy-Weinberg equilibrium. Associations between the individual tagSNPs and cardiorespiratory fitness phenotypes were analyzed using a variance components and likelihood ratio test-based procedure in the quantitative trait disequilibrium test (QTDT) software package (1). The total association model of QTDT software utilizes a variance-components framework to combine the phenotypic means model and the estimates of additive genetic, residual genetic, and residual environmental variances from a variance-covariance matrix into a single likelihood model. The evidence of association is evaluated by maximizing the likelihoods under two conditions: the null hypothesis (L_0) restricts the additive genetic effect of the marker locus to zero ($\beta_a = 0$), whereas the alternative hypothesis does not impose restrictions on β_a. The quantity of twice the difference of the log likelihoods between the alternative and the null hypotheses $\{2[\ln(L_1) - \ln(L_0)]\}$ is distributed as χ^2 with 1 degree of freedom (difference in number of parameters estimated). $\dot{V}O_{2max}$ training responses were adjusted for age, sex, baseline body weight, and baseline value of $\dot{V}O_{2max}$. Differences in allele and genotype frequencies between top and bottom quartiles of $\dot{V}O_{2max}$ training response distribution (defined using sex- and generation-specific quartile cutoffs) were tested using the case-control procedure (Proc Casecontrol) of the SAS version 9.1 Statistical Software package. Finally, the total contribution of the SNPs on $\dot{V}O_{2max}$ training response was tested using multivariate regression analysis. A backward-elimination process was used to filter out redundant SNPs due to strong pairwise LD. Then the SNPs retained by the backward-elimination model were analyzed using a stepwise regression model.

RESULTS

Generating a robust molecular profile of aerobic-training adaptation. We established a global muscle RNA expression profile in vastus lateralis in *group 1* (young sedentary) subjects

before and after 6 wk of intense aerobic exercise training. Preintervention $\dot{V}O_{2max}$ was 3.7 ± 0.6 l/min. After training, $\dot{V}O_{2max}$ increased on average by ∼14% (range −2.8% to 27.5%), while exercise respiratory exchange ratio decreased by ∼10% (at 70% of preexercise $\dot{V}O_{2max}$), indicating that we had a greater reliance on lipid substrates, as previously reported (54) (Fig. 2, *A* and *B*). With use of routinely measured physiological and biochemical biomarkers, it was not possible to predict whether an individual has a high or low potential for improved maximal aerobic capacity (54). Using significance analysis of microarrays (SAM) (52), 1,000 probe sets representing ∼800 genes (FDR <5%, >1.5-fold change) were differentially regulated during exercise training-induced aerobic adaptation [referred to hereafter as the training-responsive transcriptome (TRT); see page 1 of supplemental data]. Gene ontology analysis demonstrated that the TRT generated using the U133 Plus 2.0 platform produced a data set with biological features very similar to those of an earlier study using the U95A-E platform (50), providing confirmation of the reliability of the technology platform and the present data set. Downregulated genes had a higher baseline mean expression than upregulated genes, an interesting observation that appears logical for a dynamically remodeling tissue undergoing a transitional phase. Gene set enrichment analysis, a statistical technique that relies on the principle that genes act in a coordinated (group) manner, and not in isolation, found an upregulation in the oxidative phosphorylation (OXPHOS) gene set (FDR = 1.1%), consistent with the increased lipid oxidation profile. However, few individual OXPHOS genes were found to be changed when examined individually, consistent

Fig. 2. *A* and *B*: aerobic capacity and substrate oxidation characteristics of young sedentary human subjects before and after 6 wk of aerobic exercise training. Maximum oxygen consumption ($\dot{V}O_{2max}$) increased 14% ($P < 0.0001$) and submaximal exercise respiratory exchange ratio (RER) decreased, indicating greater reliance on lipid oxidation in vivo ($P < 0.0001$). *Insets*: geometric means ± SE pre- and posttraining. *C* and *D*: molecular predictor for aerobic adaptation to exercise training in humans involving 29 Affymetrix U133 Plus 2.0 probe sets. Logit-transformed normalized probe set intensities are correlated with ml/min increase in aerobic capacity ($\dot{V}O_{2max}$). *C*: green indicates higher gene expression, while red indicates lower gene expression. *D*: linear relationship between the sum of the expression of the molecular predictor genes and changes in aerobic capacity with endurance training derived from young sedentary subjects following 6 wk of constant-load aerobic exercise training ($n = 24$, $r^2 = 0.58$, $P < 0.00001$).
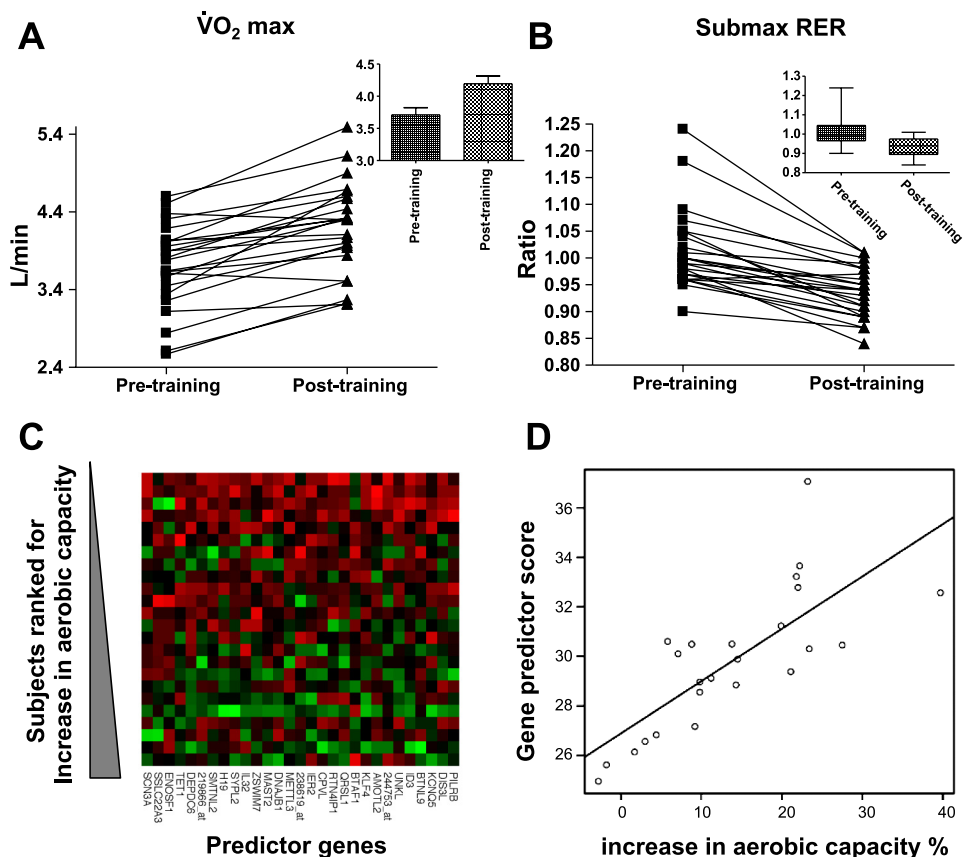
with the idea that, unlike mice, OXPHOS genes are not highly regulated in vivo in human skeletal muscle (51). The data set demonstrates that skeletal muscle adaptation reflects a distributed and a coordinated enhancement in gene networks. The TRT represents the molecular signature of ongoing and long-term aerobic adaptation, because <5% of these genes are acutely influenced by a single bout of endurance training (27, 36).

Development and validation of a quantitative predictor of response to training. With gain in aerobic capacity ($\Delta\dot{V}O_{2max}$) as our response parameter, a quantitative molecular classifier (predictor) was developed by relating $\Delta\dot{V}O_{2max}$ to baseline muscle gene RNA expression in *group 1* subjects. This analysis was repeated using a leave-one-out strategy and identified 29 significant Affymetrix probe sets (predictor genes; Fig. 2*C*). The sum of the RNA expression values of these 29 genes ($\Sigma29_{predict-RNA}$), each of which ranked highly ≥22 times during the correlation analysis, performed best in predicting $\Delta\dot{V}O_{2max}$ in *group 1* subjects (Fig. 2*D*). No further optimization was attempted. The predictor genes formed a high-scoring INGENUITY network (score = 46), characterized as "development-related" ($P < 0.001$).

To provide validation that this molecular classifier would predict the plasticity of aerobic capacity under diverse circumstances, it was tested in a blinded manner in an independent study (*group 2*). Affymetrix profiles were generated from pretraining muscle biopsy samples taken from *group 2* (preintervention $\dot{V}O_{2max}$ = 4.1 ± 0.5 l/min). These young physically active subjects underwent 10 wk of intense interval-based aerobic-training, increasing their $\dot{V}O_{2max}$, on average, by the same amount as *group 1* (~14%) and with the same heterogeneity. The $\Sigma29_{predict-RNA}$ significantly correlated to the change in $\dot{V}O_{2max}$ in the blind validation group (see supplemental Fig. S1a). Thus we demonstrated that the molecular classifier can be applied to human subjects with a wide range in aerobic fitness capacities and confirmed the validity of the gene selection process. The $\Sigma29_{predict-RNA}$ genes were then used to assist in the identification of the key genetic variants predicting the ability to alter aerobic capacity in response to exercise training.

Validation of molecular classifier using genetic association analysis. A new analysis of the HERITAGE Family Study ($n = 473$) was carried out using ~300 tagSNPs for the 29 predictor probe sets. Sedentary subjects from 99 nuclear families were trained for 20 wk with a fully standardized and monitored exercise program. The mean gain in $\dot{V}O_{2max}$ was similar to that in the studies described above (~400 ml O_2), with a standard deviation of ~200 ml O_2. With use of a model-fitting procedure, the heritability of the changes in $\dot{V}O_{2max}$ was calculated to be 47% (7), and thus our genotyping analysis could, at most, expect to capture ~47% of the total variance in the gains in maximal aerobic capacity. We were able to identify six genes from the predictor signature set that harbored genetic variants associated with gains in aerobic capacity ($P < 0.01$ for each). When the upper quartile of the $\dot{V}O_{2max}$ response distribution was compared with the lower quartile, SNPs in *SMTNL2, DEPDC6, SLC22A3, METTL3,* and *BTNL9* differed in genotype or allele frequencies (see supplemental data). In addition, in the comparison of the $\dot{V}O_{2max}$ response by genotype for the entire HERITAGE Family Study population, a variant in *ID3* was evidenced (rs11574; $P = 0.0058$). With use of all relevant genetic variants identified (see supplemental Table S4) from the molecular predictor ($n = 25$) plus those from ongoing quantitative trait locus (QTL) and candidate gene studies within the HERITAGE project ($n = 10$), a stepwise regression model was applied using the residual $\dot{V}O_{2max}$ responses, adjusted for major confounding variables. The results were striking: 11 SNPs captured 23% of the total variance in aerobic capacity responses (Table 1). Reciprocal analysis, where HERITAGE-derived genes retained in the model were tested in the RNA expression model, independently validated three of four of the HERITAGE genes. Furthermore, addition of *SVIL* and *NRP2* yielded an improved correlation coefficient ($r^2 = 0.60$) and/or stronger P value ($P = 0.009$) for the RNA-based blind validation data set (*group 2*; see supplemental Fig. S2b). Interestingly, *MIPEP* RNA expression was negatively correlated ($r^2 = -0.64$, $P = 0.0051$) but did not worsen or improve the performance of the tissue-based classifiers.

Table 1. *Stepwise regression model for standardized residuals of maximal oxygen consumption training response in the HERITAGE Family Study*

Gene (SNP)	Identification Method	RNA-Level Correlation	RNA Stable to Exercise	Genomic Location	r^2 Partial	r^2 Model	P
SVIL (rs6481619)	QTL	Yes (+)	Yes	10p11.2	0.0411	0.0411	<.0001
SLC22A3 (rs2457571)	RNA predictor	Yes (+)	Yes	6q26-q27	0.0307	0.0718	0.0003
NRP2 (rs3770991)	QTL	Yes (+)	Yes	2q33.3	0.0224	0.0942	0.0017
TTN (rs10497520)	QTL	No	Yes	2q31	0.0204	0.1146	0.0025
H19 (rs2251375)	RNA predictor	Yes (+)	No	11p15.5	0.0268	0.1414	0.0004
ID3 (rs11574)	RNA predictor	Yes (+)	Yes	1p36.13-p36.12	0.02	0.1615	0.0021
MIPEP (rs7324557)	QTL	Yes (−)	Yes	13q12	0.0163	0.1778	0.0051
CPVL (rs4257918)	RNA predictor	Yes (+)	Yes	7p15-p14	0.0179	0.1957	0.0031
DEPDC6 (rs7386139)	RNA predictor	Yes (+)	Yes	8q24.12	0.0112	0.2069	0.0185
BTAF1 (rs2792022)	RNA predictor	Yes (+)	Yes	10q22-q23	0.0125	0.2194	0.0122
DIS3L (rs1546570)	RNA predictor	Yes (+)	Yes	15q22.31	0.0095	0.2289	0.0279

Residual maximal oxygen consumption response values were adjusted for age, sex, baseline body weight, and baseline maximal oxygen consumption ($n = 473$). RNA determinations were made using U133 Plus 2.0 Affymetrix chips, normalized using MAS5.0, and tested for differential expression using significance analysis of microarrays (SAM). In total, ~1,000 probe sets were deemed differentially expressed (>1.5-fold change, <5% false discovery rate). Only 1 predictor/genetic association gene belonged to this group of regulated genes. SNP, single-nucleotide polymorphism; QTL, quantitative trait locus.

DISCUSSION

It is reasonable to state that molecular classifiers (predictors) will be essential for implementing personalized medicine, yet there are limited examples of validated predictors that are able to tailor interventions relevant to the most pressing factors impacting on public health. Rather, those that have been discovered typically relate to rarer single-gene polymorphisms and selective drug toxicity. In the present study, we produce and validate a molecular classifier that tracks a much more complex phenotype, i.e., aerobic capacity changes in humans subjected to endurance exercise training. While the RNA predictor represents a significant development in its own right, generating the RNA predictor also led us (by identifying candidate genes for genotyping) to discover ~50% of the estimated genetic variance for the $\dot{V}_{O_{2max}}$ gains following aerobic training in humans. The outcome of our study suggests that using RNA-based molecular classification to direct targeted genotyping can be a powerful and new strategy for studying the heterogeneous responses of additional physiological phenotypes.

Regulators of cardiovascular adaptation to endurance training. Maximal aerobic power is arguably one of the most important physiological properties that can be manipulated to prevent cardiovascular disease (55, 57). An association between maximal oxidative metabolism and health may be linked, to some extent, by the notion that evolution of multicellular organisms was afforded by the steep thermodynamic gradient of an oxygen-enriched atmosphere (4, 43). Under this scenario, evolving biochemical strategies, aimed at coping with the obligatory reactive nature of oxygen chemistry (47), would have firmly shaped our genome in a manner that is still obscure. In the present study, an important observation is that the molecular classifier genes have been associated with tissue development, oxidative gene activation, and imprinting (see below), implying that aerobic exercise adaptability may be reliant on or predetermined by molecular mechanisms responsible for regulating organogenesis and angiogenesis, in particular, mechanisms clearly linked to the sensing and delivery of molecular oxygen. Indeed, developmental pathways have been associated with cardiovascular disease (40) and metabolic syndrome (17), providing support for the likelihood that this association could be informative.

Much consideration has been given to the biochemical signals that drive adaptation in response to exercise training, with the aim to better understand the health benefits of physical activity. Mutation and knockout of candidate genes, such as AMP-activated protein kinase (AMPK) and peroxisome proliferator-activated receptor-γ coactivator-1α (PGC-1α), have been postulated as causal regulators of muscle adaptation (2, 21, 26, 31). However, these genes do not appear critical for exercise-induced endurance-training adaptation in terms of muscle biochemistry or performance (2, 21, 26, 31), also suggesting that single-gene knockouts do not necessarily facilitate physiological understanding. In humans, when a global analysis was carried out on the TRT in this study or previous studies (23, 24, 27, 42, 50), there was little evidence generated suggesting that AMPK or PGC-1α links to the gene networks that characterize muscle phenotype changes occurring with endurance training (e.g., gene set enrichment analysis did not identify any PGC-1α-related transcription factors in the TRT).

Furthermore, we could not demonstrate a link between metabolic homeostasis, directly measured in vivo in human muscle for the first time at the training intensity, and subsequent changes in aerobic capacity (54).

There have been many acute exercise canonical pathway signaling studies in humans (3, 22, 41, 56) also aimed at discovering the molecules responsible for "translating" acute exercise effects into long-term exercise adaptation. However, a pathway analysis of the acute response to endurance exercise (36) appears to suggest that the immediate consequences reflect molecular responses to short-term energy and ionic homeostasis challenges. Ontological analysis of the TRT (see supplemental data) supports the idea that extracellular matrix and major histocompatability complex modulation and calcium signaling may be the potential transducers of cardiovascular adaptation in humans, connecting mechanical work to muscle tissue adaptation during endurance training. These observations suggest that acutely regulated molecules, such as AMPK and IL-6, are not related to the adaptive process. Rather, modulation of these molecules may be more related to recovery of short-term energy homeostasis within the muscle or reflect a deregulation of muscle molecular phenotype when the cell is exposed to an acute energy "crisis." A further point for consideration is that, given the very large range of training-induced adaptation noted in humans (34), one would need to establish, in each acute exercise study, whether the responses measured following acute exercise actually lead to, or associate with, a measure of chronic adaptation of the physiological system being studied. Indeed, it is very clear that many of the ~800 TRT genes are equally regulated in high and low responders, an analysis that highlights the idea that only a subset of genes may "set" or limit the potential for physiological adaptation. However, rarely, for any physiological system, will this "minimum" set be fewer than two genes, and thus many "key" genes will be required to coordinate adaptation.

Do the predictors of cardiovascular adaptation to endurance training help identify causal regulators? The present study highlights an alternative approach to transgenic models for producing candidate mechanisms for subsequent hypothesis testing, an approach that is reliant on in vivo data generated under conditions relevant for human disease and physiology. For example, the biological profile of the TRT is robust (reproduced across different chip generations and biological samples) and provides new candidates that regulate training adaptation. However, it is challenging to establish a direct cause-effect relationship between these ~800 RNA changes and physiological adaptation. In contrast, the evidence for a causal role for the 29 predictor genes and, in particular, the 11 SNP genes that captured ~50% of the estimated variance attributable to genetic differences can be considered much stronger and, thus, worthy of some detailed discussion.

First, the genotype-RNA associations appear to be driven by factors (including genetic variation) independent of chronic exercise training, as the RNA expression levels of the predictor genes (which capture ~50% of the total variance for changes in maximal aerobic capacity) were stable to exercise training. Furthermore, the RNA expression was remarkably stable to exercise in low and high responders (Fig. 3). Thus the largely higher "preset" levels of the 29 predictor genes appear to offer high responders a greater potential to activate some components of the TRT. While the vast majority of the genes within
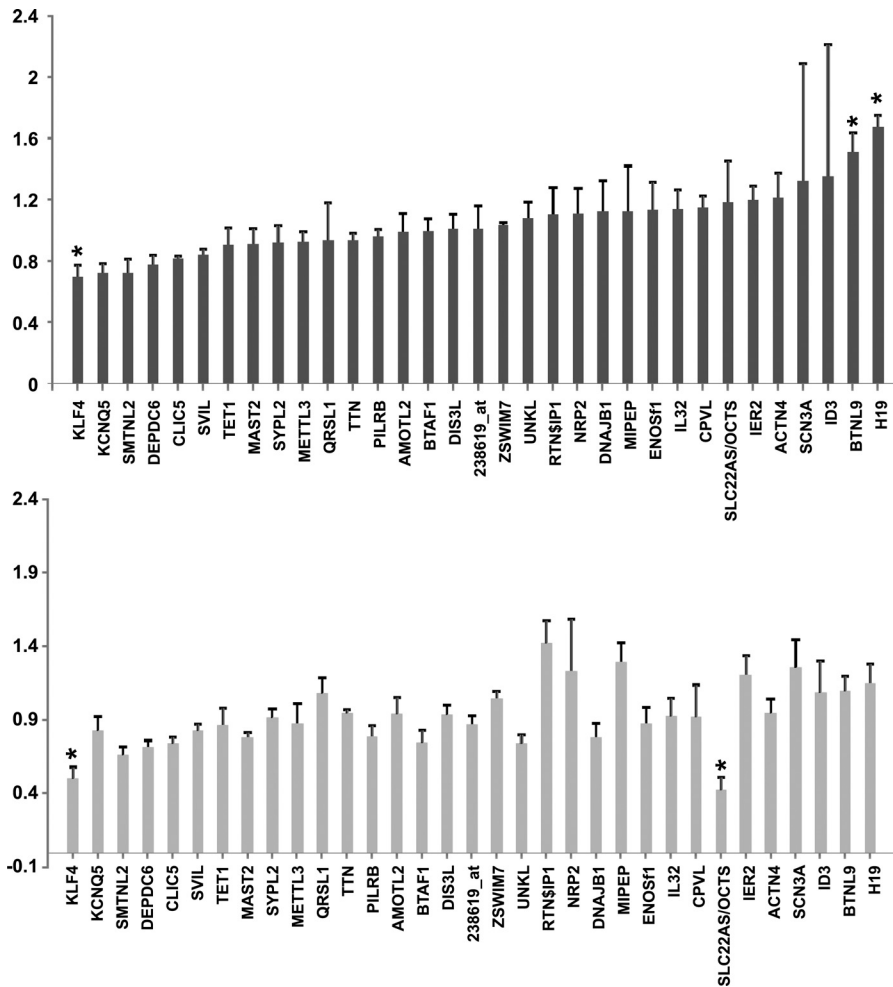
Fig. 3. Skeletal muscle expression of all the predictor genes (based on RNA), which together explain ~58% of the variance in exercise training-induced changes in $\dot{V}O_{2max}$ in young sedentary human subjects ($n = 24$) before and after 6 wk of aerobic exercise training. *Top*: high responders following 6 wk of aerobic training; *bottom*: low responders following 6 wk of aerobic training. Expression values are generated using Affymetrix U133 Plus 2.0 gene chips (>47,000 transcripts), normalized using MAS5.0, and corrected for multiple comparisons using the significance of microarrays analysis methodology. This analysis strategy avoids issues with unstable housekeeping genes. Only *H19, KLF4, OCT3, SMTNL2,* and *BTNL9* demonstrated a modest change in expression below a false discovery rate of 5% (∗), and not consistently in high and low responders. All other genes (>90% of predictor genes) were unchanged.

the predictor gene set are not well studied, especially in the context of cardiovascular biology, some do provide obvious links to cardiovascular adaptation and proposed signaling molecules involved in cardiovascular adaptation and cardiovascular disease (Fig. 4A). Network analysis provides additional linking molecules, some of which are regulated by exercise (Fig. 4B). For example, *ID3* is a *TGFβ1*- and superoxide-regulated gene that interacts (39) with another member of the baseline predictor (*KLF4*) and appears essential for angiogenesis (35). Transforming growth factor-β (TGFβ) signaling is important for any tissue remodeling, and, clearly, genes that alter the potential for a safe and controlled proangiogenesis are highly relevant for expansion of maximal oxygen transport capacity.

In addition, the imprinted transcript *SLC22A3* (*OCT3*), which harbored genetic variation associated with training response ($P = 0.0047$), is part of the Air noncoding RNA imprinted locus mechanism, which interacts (38) with another of our predictor genes, *H19*. This suggests that our predictor genes may participate in the regulation of imprinting and that the mechanisms that link aerobic capacity and cardiovascular-metabolic disease may share common features with developmental processes (18, 53). Interestingly, *OCT3* and *H19* were the predictor genes most differentially regulated between high and low responders (Fig. 3). *OCT3* RNA appears downregulated in only high responders, while *H19* RNA is unchanged. In

contrast, *OCT3* RNA is unchanged in the low responders, while *H19* RNA is significantly upregulated (FDR <5%). Whether this relates to the longer-term mechanisms operational during imprinting and, hence, regulation of insulin-like growth factor II (*IGF2*) signaling is unclear but merits detailed analysis (Fig. 4B). However, we would still caution against the overinterpretation that this discrete list of genes alone determines aerobic adaptation but, rather, prefer to assume that they appear to be important and may directly or indirectly influence many of the genes we show to be regulated by exercise (e.g., TRT). Briefly, while we can generate a molecular classifier for the purpose of diagnosis or for randomization in future hypothesis-testing studies, it is unclear whether any gene can be ruled out from playing at least a minor role in the complex in vivo process of aerobic adaptation, even the much emphasized AMPK, PGC-1α, and IL-6 systems.

Conclusions and Future Perspectives

Despite the use of baseline physiological data to personalize the clinical intervention in all three studies, ~20% of subjects demonstrated <5% improvement in maximal aerobic capacity. It should be noted that our molecular screen for exercise responsiveness worked across diverse situations (short- and long-term moderate-intensity aerobic-training and interval-based maximal exercise-training regimens), such that failure to
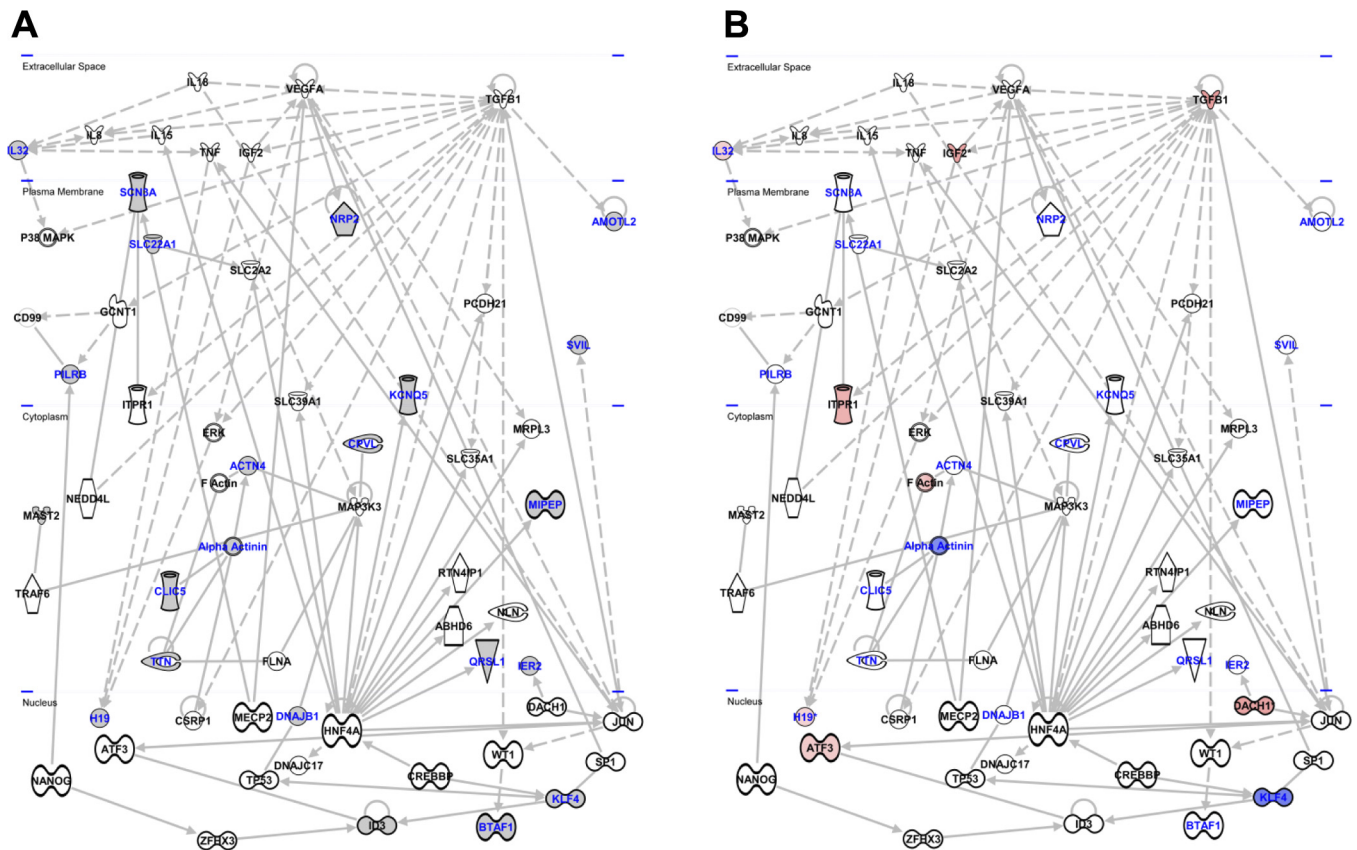
Fig. 4. Genes identified as being able to predict which subjects demonstrate high response to exercise training formed a gene network found to be related to development. *A*: predictor genes (gray symbols with blue labeling) are not regulated by exercise (Table 1). Additional genes within the network represent genes known to interact with the predictor genes. *B*: genes shown in red are those that are increased by exercise training in *group 1* (while those shown in blue are decreased). This demonstrates that most of the predictor genes, and many of those with which they interact in this analysis, are not responsive to regular exercise. Several members of the network are, however, known to be regulated by exercise at the protein level.

adapt to exercise is a generalized observation and is not model-specific. The fact that we can forecast gains in aerobic capacity using a peripheral gene signature (validated using 2 independent clinical cohorts) demonstrates that RNA profiling captures critical information and is able to inform about an integrated physiological processes. If we consider the strong relationships between aerobic capacity and morbidity and mortality, the approach defined in this study has the potential to provide insight into the role of regular exercise and its molecular determinants in metabolic control of glucose and lipid metabolism, blood pressure regulation, and inflammation in humans.

Given that ~20% of subjects fail to improve aerobic capacity with intense endurance training, while ~30% do not enhance their insulin sensitivity (12), it is clear that we must develop genotype-tailored lifestyle interventions. The present data enable us to test the efficacy of a molecular screen for exercise responses (8), so that, for example, low responders for $\dot{V}_{O_{2max}}$ can be randomized to alternative interventions to establish if they are low responders for additional parameters, such as changes in postexercise insulin sensitivity. Such an application of our data is consistent with the idea of using personalized medicine (48) to tailor lifestyle modification, and, hence, we believe that we provide a significant new milestone for the field. For those individuals with an impaired ability to improve or even maintain their aerobic capacity with exercise

therapy, we may need to promote alternative exercise intervention paradigms or offer more intensive pharmacological and dietary protocols to help compensate for their genomic profile. Identification of individuals with a particular risk profile is the first step toward the implementation of personalized medicine.

DISCLOSURES

A patent has been filed for the low-responder predictor.

REFERENCES

1. **Abecasis GR, Cardon LR, Cookson WO.** A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 66: 279–292, 2000.

2. **Akimoto T, Ribar TJ, Williams RS, Yan Z.** Skeletal muscle adaptation in response to voluntary running in Ca^{2+}/calmodulin-dependent protein kinase IV-deficient mice. *Am J Physiol Cell Physiol* 287: C1311–C1319, 2004.

3. **Baar K, Wende AR, Jones TE, Marison M, Nolte LA, Chen M, Kelly DP, Holloszy JO.** Adaptations of skeletal muscle to exercise: rapid increase in the transcriptional coactivator PGC-1. *FASEB J* 16: 1879–1886, 2002.

4. **Baldwin JE, Krebs H.** The evolution of metabolic cycles. *Nature* 291: 381–382, 1981.

5. **Blair SN, Kampert JB, Kohl HW 3rd, Barlow CE, Macera CA, Paffenbarger RS Jr, Gibbons LW.** Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. *JAMA* 276: 205–210, 1996.

6. **Blair SN, Kohl HW 3rd, Paffenbarger RS Jr, Clark DG, Cooper KH, Gibbons LW.** Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA* 262: 2395–2401, 1989.

7. **Bouchard C, An P, Rice T, Skinner JS, Wilmore JH, Gagnon J, Perusse L, Leon AS, Rao DC.** Familial aggregation of $\dot{V}O_{2max}$ response to exercise training: results from the HERITAGE Family Study. *J Appl Physiol* 87: 1003–1008, 1999.

8. **Bouchard C, Chagnon M, Thibault MC, Boulay MR, Marcotte M, Cote C, Simoneau JA.** Muscle genetic variants and relationship with performance and trainability. *Med Sci Sports Exerc* 21: 71–77, 1989.

9. **Bouchard C, Leon AS, Rao DC, Skinner JS, Wilmore JH, Gagnon J.** The HERITAGE Family Study. Aims, design, and measurement protocol. *Med Sci Sports Exerc* 27: 721–729, 1995.

10. **Bouchard C, Rankinen T.** Individual differences in response to regular physical activity. *Med Sci Sports Exerc* 33: S446–S453, 2001.

11. **Bouchard C, Rankinen T, Chagnon YC, Rice T, Perusse L, Gagnon J, Borecki I, An P, Leon AS, Skinner JS, Wilmore JH, Province M, Rao DC.** Genomic scan for maximal oxygen uptake and its response to training in the HERITAGE Family Study. *J Appl Physiol* 88: 551–559, 2000.

12. **Boule NG, Weisnagel SJ, Lakka TA, Tremblay A, Bergman RN, Rankinen T, Leon AS, Skinner JS, Wilmore JH, Rao DC, Bouchard C.** Effects of exercise training on glucose homeostasis: the HERITAGE Family Study. *Diabetes Care* 28: 108–114, 2005.

13. **Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, Leonardson A, Castellini LW, Wang S, Champy MF, Zhang B, Emilsson V, Doss S, Ghazalpour A, Horvath S, Drake TA, Lusis AJ, Schadt EE.** Variations in DNA elucidate molecular networks that cause disease. *Nature* 452: 429–435, 2008.

14. **Choe SE, Boutros M, Michelson AM, Church GM, Halfon MS.** Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. *Genome Biol* 6: R16, 2005.

15. **Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA.** DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4: P3, 2003.

16. **Frazer KA, Murray SS, Schork NJ, Topol EJ.** Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 10: 241–251, 2009.

17. **Gesta S, Bluher M, Yamamoto Y, Norris AW, Berndt J, Kralisch S, Boucher J, Lewis C, Kahn CR.** Evidence for a role of developmental genes in the origin of obesity and body fat distribution. *Proc Natl Acad Sci USA* 103: 6676–6681, 2006.

18. **Gluckman PD, Hanson MA.** Developmental plasticity and human disease: research directions. *J Intern Med* 261: 461–471, 2007.

19. **Gulati M, Pandey DK, Arnsdorf MF, Lauderdale DS, Thisted RA, Wicklund RH, Al-Hani AJ, Black HR.** Exercise capacity and the risk of death in women: the St. James Women Take Heart Project. *Circulation* 108: 1554–1559, 2003.

20. **Hamilton MT, Booth FW.** Skeletal muscle adaptation to exercise: a century of progress. *J Appl Physiol* 88: 327–331, 2000.

21. **Handschin C, Spiegelman BM.** The role of exercise and $PGC1\alpha$ in inflammation and chronic disease. *Nature* 454: 463–469, 2008.

22. **Hawley JA, Hargreaves M, Zierath JR.** Signalling mechanisms in skeletal muscle: role in substrate selection and muscle adaptation. *Essays Biochem* 42: 1–12, 2006.

23. **Hittel DS, Kraus WE, Hoffman EP.** Skeletal muscle dictates the fibrinolytic state after exercise training in overweight men with characteristics of metabolic syndrome. *J Physiol* 548: 401–410, 2003.

24. **Hittel DS, Kraus WE, Tanner CJ, Houmard JA, Hoffman EP.** Exercise training increases electron and substrate shuttling proteins in muscle of overweight men and women with the metabolic syndrome. *J Appl Physiol* 98: 168–179, 2005.

25. **Hosack DA, Dennis G Jr, Sherman BT, Lane HC, Lempicki RA.** Identifying biological themes within lists of genes with EASE. *Genome Biol* 4: R70, 2003.

26. **Jorgensen SB, Jensen TE, Richter EA.** Role of AMPK in skeletal muscle gene adaptation in relation to exercise. *Appl Physiol Nutr Metab* 32: 904–911, 2007.

27. **Keller P, Vollaard NJB, Babraj J, Ball D, Sewell DA, Timmons JA.** Using systems biology to define the essential biological networks responsible for adaptation to endurance exercise training. *Biochem Soc Trans* 35: 1306–1309, 2007.

28. **Knudsen S, Knudsen S.** *Guide to Analysis of DNA Microarray Data.* Hoboken, NJ: Wiley-Liss, 2004, p. xv.

29. **Kokkinos P, Myers J, Kokkinos JP, Pittaras A, Narayan P, Manolis A, Karasik P, Greenberg M, Papademetriou V, Singh S.** Exercise capacity and mortality in black and white men. *Circulation* 117: 614–622, 2008.

30. **Larsson O, Wahlestedt C, Timmons JA.** Considerations when using the significance analysis of microarrays (SAM) algorithm. *BMC Bioinformatics* 6: 129, 2005.

31. **Leick L, Wojtaszewski JF, Johansen ST, Kiilerich K, Comes G, Hellsten Y, Hidalgo J, Pilegaard H.** $PGC-1\alpha$ is not mandatory for exercise- and training-induced adaptive gene responses in mouse skeletal muscle. *Am J Physiol Endocrinol Metab* 294: E463–E474, 2008.

32. **Li C, Wong WH.** Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol* 2: RESEARCH0032, 2001.

33. **Li C, Wong WH.** Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci USA* 98: 31–36, 2001.

34. **Lortie G, Simoneau JA, Hamel P, Boulay MR, Landry F, Bouchard C.** Responses of maximal aerobic power and capacity to aerobic training. *Int J Sports Med* 5: 232–236, 1984.

35. **Lyden D, Young AZ, Zagzag D, Yan W, Gerald W, O'Reilly R, Bader BL, Hynes RO, Zhuang Y, Manova K, Benezra R.** Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. *Nature* 401: 670–677, 1999.

36. **Mahoney DJ, Parise G, Melov S, Safdar A, Tarnopolsky MA.** Analysis of global mRNA expression in human skeletal muscle during recovery from endurance exercise. *FASEB J* 19: 1498–1500, 2005.

37. **Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood JE.** Exercise capacity and mortality among men referred for exercise testing. *N Engl J Med* 346: 793–801, 2002.

38. **Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P.** The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322: 1717–1720, 2008.

39. **Nickenig G, Baudler S, Muller C, Werner C, Werner N, Welzel H, Strehlow K, Bohm M.** Redox-sensitive vascular smooth muscle cell proliferation is mediated by GKLF and Id3 in vitro and in vivo. *FASEB J* 16: 1077–1086, 2002.

40. **Olivotto I, Cecchi F, Poggesi C, Yacoub MH.** Developmental origins of hypertrophic cardiomyopathy phenotypes: a unifying hypothesis. *Nat Rev Cardiol* 6: 317–321, 2009.

41. **Pedersen BK, Febbraio MA.** Muscle as an endocrine organ: focus on muscle-derived interleukin-6. *Physiol Rev* 88: 1379–1406, 2008.

42. **Radom-Aizik S, Hayek S, Shahar I, Rechavi G, Kaminski N, Ben-Dov I.** Effects of aerobic training on gene expression in skeletal muscle of elderly men. *Med Sci Sports Exerc* 37: 1680–1696, 2005.

43. **Raymond J, Segre D.** The effect of oxygen on biochemical networks and the evolution of complex life. *Science* 311: 1764–1767, 2006.

44. **Saltin B, Calbet JA.** Point: in health and in a normoxic environment, $\dot{V}O_{2max}$ is limited primarily by cardiac output and locomotor muscle blood flow. *J Appl Physiol* 100: 744–745, 2006.

45. **Saxena R, de Bakker PI, Singer K, Mootha V, Burtt N, Hirschhorn JN, Gaudet D, Isomaa B, Daly MJ, Groop L, Ardlie KG, Altshuler D.** Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. *Am J Hum Genet* 79: 54–61, 2006.

46. **Snyder M, Weissman S, Gerstein M.** Personal phenotypes to go with personal genomes. *Mol Syst Biol* 5: 273, 2009.

47. **Strohman R.** Thermodynamics—old laws in medicine and complex disease. *Nat Biotechnol* 21: 477–479, 2003.

48. **Timmons JA.** Perspective on the future use of genomics in exercise prescription. *J Appl Physiol* 104: 1250, 2008.

49. **Timmons JA, Gustafsson T, Sundberg CJ, Jansson E, Greenhaff PL.** Muscle acetyl group availability is a major determinant of oxygen deficit in humans during submaximal exercise. *Am J Physiol Endocrinol Metab* 274: E377–E380, 1998.

50. **Timmons JA, Larsson O, Jansson E, Fischer H, Gustafsson T, Greenhaff PL, Ridden J, Rachman J, Peyrard-Janvid M, Wahlestedt C, Sundberg CJ.** Human muscle gene expression responses to endurance training provide a novel perspective on Duchenne muscular dystrophy. *FASEB J* 19: 750–760, 2005.

51. **Timmons JA, Norrbom J, Scheele C, Thonberg H, Wahlestedt C, Tesch P.** Expression profiling following local muscle inactivity in humans provides new perspective on diabetes-related genes. *Genomics* 87: 165–172, 2006.

52. **Tusher VG, Tibshirani R, Chu G.** Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 98: 5116–5121, 2001.

53. **van Hoek M, Langendonk JG, de Rooij SR, Sijbrands EJ, Roseboom TJ.** A genetic variant in the IGF2BP2 gene may interact with fetal malnutrition on glucose metabolism. *Diabetes* 58: 1440–1444, 2009.

54. **Vollaard NB, Constantin-Teodosiu D, Fredriksson K, Rooyackers OE, Jansson E, Greenhaff PL, Timmons JA, Sundberg CJ.** Systematic analysis of adaptations in aerobic capacity and submaximal energy metabolism provides a unique insight into determinants of human aerobic performance. *J Appl Physiol* 106: 1479–1486, 2009.

55. **Wei M, Kampert JB, Barlow CE, Nichaman MZ, Gibbons LW, Paffenbarger RS Jr, Blair SN.** Relationship between low cardiorespiratory fitness and mortality in normal-weight, overweight, and obese men. *JAMA* 282: 1547–1553, 1999.

56. **Widegren U, Jiang XJ, Krook A, Chibalin AV, Bjornholm M, Tally M, Roth RA, Henriksson J, Wallberg-henriksson H, Zierath JR.** Divergent effects of exercise on metabolic and mitogenic signaling pathways in human skeletal muscle. *FASEB J* 12: 1379–1389, 1998.

57. **Wisloff U, Najjar SM, Ellingsen O, Haram PM, Swoap S, Al-Share Q, Fernstrom M, Rezaei K, Lee SJ, Koch LG, Britton SL.** Cardiovascular risk factors emerge after artificial selection for low aerobic capacity. *Science* 307: 418–420, 2005.