Natural language indicators of differential gene regulation in the human immune system

Matthias Mehl[1], Charles Raison[2], Thaddeus Pace[1], Jesusa Arevalo[3], Steven W. Cole[4]

[1]University of Arizona, [2]University of Wisconsin-Madison, [3]UCLA School of Medicine, [4]UCLA

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Adverse social conditions have been linked to a conserved transcriptional response to adversity (CTRA) in circulating leukocytes that may contribute to social gradients in disease. However, the central nervous system (CNS) mechanisms involved remain obscure in part because CTRA gene expression profiles often track external social-environmental variables more closely than they do self-reported internal affective states such as stress, depression, or anxiety. This study examined the possibility that variations in patterns of natural language use might provide more sensitive indicators of the automatic threat detection and response systems that proximally regulate autonomic induction of the CTRA. In 22,627 audio samples of natural speech sampled from the daily interactions of 143 healthy adults, both total language output and patterns of function word use covaried with CTRA gene expression. These language features predicted CTRA gene expression substantially better than did conventional self-report measures of stress, depression, and anxiety, and did so independently of demographic and behavioral factors (age, sex, race, smoking, BMI) and leukocyte subset distributions. Prediction held when language and gene expression were sampled more than a week apart, suggesting that associations reflect stable individual differences or chronic life circumstances. Given the observed relationship between personal expression and gene expression, patterns of natural language use may provide a useful behavioral indicator of non-consciously evaluated well-being (implicit safety vs. threat) that is both distinct from conscious affective experience and more closely tracks the neurobiological processes involved in peripheral gene regulation.

genomics | psychoneuroimmunology | psycholinguistics

Across a diverse array of adverse life circumstances such as low socioeconomic status, social isolation, diagnosis with a life-threatening disease, and post-traumatic stress, circulating immune cells have been found to show a conserved transcriptional response to adversity (CTRA) marked by up-regulated expression of pro-inflammatory genes and down-regulated expression of genes involved in Type I interferon antiviral responses and IgG antibody synthesis (1, 2). These effects are mediated peripherally by sympathetic nervous system (SNS) activation of β-adrenergic signaling pathways that regulate gene transcription in existing cells and stimulate hematopoietic development of new myeloid lineage leukocytes (particularly monocytes) (3, 4). Peripheral SNS activity is controlled by a central network of brain structures including the insular and anterior cingulate cortex, extended amygdala, and lateral hypothalamus, which collectively regulate sympathetic outflow from the medulla oblongata (5, 6). Despite the clear role of central nervous system (CNS) processes in regulating SNS activity, CTRA gene expression has shown less consistent association with self-report measures of internal affective experience (e.g., stress, depression, or negative emotions) than it has with measures of external social conditions (e.g., socio-economic status, social rank, bereavement) or subjective perceptions of those external conditions (e.g., perceived isolation) (7-11). Adverse environments are clearly "getting under the skin" somehow, and this presumably involves some form of CNS information processing (as neuropharmacologic and behavioral interventions can inhibit the CTRA (3, 12, 13)), but the specific psychological processes involved remain poorly understood. In the present research we sought to identify an objectively observable behavioral indicator of the psychological processes involved in generating stable individual differences in basal CTRA gene expression.

Findings from affective neuroscience suggest that conscious experiences of negative emotional states such as stress, fear, or anxiety are mediated by a neocortical system which is functionally distinct from the more basic automatic threat detection and response system that proximally regulates SNS activity (see (6) for more detail on functional and neuroanatomical distinction between the conscious fear system and the automatic threat defense system). If the non-conscious threat defense system is more sensitive to adverse social conditions than is the conscious affect system, that could explain why CTRA gene expression more reliably tracks measures of adverse environmental conditions than it does self-report measures of experienced affect (Fig S1). However, this raises a significant methodological challenge for efforts to map the central psychological processes that mediate "social signal transduction" (1, 2): if psychometric self-report instruments cannot accurately track the activity of the relevant neural system, then how instead should we measure it?

Recent psycholinguistic analyses have found that patterns of natural language use change systematically under threatening conditions such as social deception (14-17), terrorist attack (18), low social status (19), and personal crisis (20). These changes include alterations in total language output and a shift in the use of specific function words such as pronouns (17). Function words (e.g., articles, adverbs, pronouns) are generated relatively automatically by CNS language systems and serve to map relationships among the more consciously generated meaning words (e.g., nouns or verbs) that carry the primary semantic informa-

Significance

Social genomics research has identified a conserved transcriptional response to adversity (CTRA) that may contribute to social disparities in health. This study identified systematic individual differences in natural language use that track CTRA gene expression more closely than do conventional self-report measures of stress, anxiety, or depression. These language style markers may provide a useful behavioral indicator of the neurobiological processes that mediate social influences on gene expression in immune cells.

Reserved for Publication Footnotes

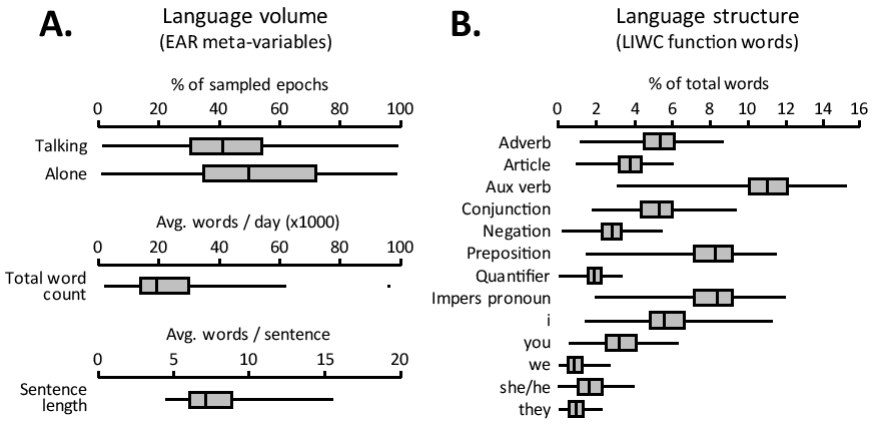

Fig. 1. Natural language use features. Distribution of individual differences in **(A)** electronically activated recorder (EAR)-sampled parameters of language volume and **(B)** language structure, including 8 general categories of function word (capitalized labels) and 5 subcategories of personal pronoun (lower case labels). Data represent average values for each parameter computed over a mean of 158 (range 14-260) 30-50 s audio samples collected at 9-12.5 min intervals over 2 days (resulting in an average 4,070 ± 272 words per individual). Whiskers indicate the range of individual values across 143 study participants, boxes span 25th-75th percentiles, and internal bars indicate 50th percentile.

tion speakers intend to convey (17, 21, 22). Function words by themselves have no semantic purpose, but together they help provide the structure of syntax. Given the relatively automatic production of function words and their empirical sensitivity to threat, systematic variations in language structure may provide an implicit behavioral indicator of the nonconscious threat response system that regulates SNS activity and, by extension, CTRA gene expression (3, 4) (see Fig S1 for a graphical depiction of this model). We tested this hypothesis using unobtrusive ecological speech sampling (23, 24) and identified systematic individual differences in broad patterns of natural language use that track CTRA gene expression better than do conventional self-report measures of stress, depression, and anxiety.

Results

Natural language use was assessed in 22,627 30-50 s audio samples collected unobtrusively at 9-12.5 min intervals over 2 days from each of 143 healthy community-dwelling adults during waking hrs. Speech samples were acquired using the Electronically Activated Recorder (EAR) system (23, 24), which generated an average 158 (range 14-260) audio samples per individual. Audio samples were coded for technical validity (i.e., participants being awake and wearing the EAR device) and the presence of others, and all valid samples were transcribed to isolate the study participant's speech. These analyses yielded individual summary values on 4 meta-variables reflecting total language volume (fraction of samples in which the individual spoke, fraction in which the individual was alone, average words spoken, and average sentence length; Fig 1a). Participant speech transcripts were subsequently processed using the Linguistic Inquiry and Word Count (LIWC) system to quantify individual differences in general language style (25). These analyses yielded 13 dimensions of language structure (Fig 1b) including prevalence of 8 general categories of function words (adverbs, articles, auxiliary verbs, conjunctions, negations, prepositions, quantifiers, and impersonal pronouns) and 5 subcategories of personal pronoun (1st person singular, 1st person plural, 2nd person singular, 3rd person singular, and 3rd person plural, each represented separately due to their potentially divergent patterns of change under threat (17)).

Characteristics of the study sample are given in Table S1 and reflect generally healthy young adults (mean age 34.2 ± SE 0.7 years; range 25-56) from the Atlanta metropolitan area with a preponderance of females (66%), multiple races (30% African American, 59% White, 7% Asian, 4% Other), and low behavioral health risk factors (mean BMI 24.3 ± 0.3 kg/m^2 with 8% BMI > 30 and 8% history of regular smoking). Self-report psychometric measures of stress, depression, anxiety, and loneliness were moderately correlated (*r*'s ranging from .44 to .64; Dataset S1) but showed no substantial association with language metrics (all $|r| < .25$; Dataset S1). Several language metrics were also correlated (Dataset S1), and the fraction of audio samples in which the subject spoke (talk frequency) showed particularly strong correlation with total word counts and frequency of being alone (both $|r| > .70$, $p < .0001$). Talk frequency was thus analyzed separately to avoid multicollinearity (variance inflation factor = 8.3). No other language metrics showed substantial multicollinearity (variance inflation factors ranged from 1.4 to 2.8, all well below the material threshold of 10 (26)).

CTRA gene expression

Expression of 50 CTRA indicator genes was assessed using microarray-based transcriptome profiling of peripheral blood mononuclear cell samples that were collected under resting conditions following the 2-day language sampling period. The CTRA profile was quantified as an a priori-defined contrast across the 50 indicator variables (inflammation-related transcripts weighted +1 and antiviral and antibody-related transcripts weighted -1 (8, 10, 27, 28)) and tested for association with 5 general domains of predictor variables: demographic and health behavior parameters (age, sex, race, BMI, and smoking), language volume measures (word count, words per sentence, frequency of being alone vs. with others), language structure measures (relative frequency of 13 function word classes), self-report psychometric measures (depression, anxiety, perceived stress, and loneliness), and RNA-based measures of leukocyte subset distribution (CD4+ and CD8+ T lymphocytes, B lymphocytes, Natural Killer cells, and monocytes). Each domain was comprised of multiple sub-dimensions which were tested simultaneously using a single omnibus partial *F* test (26) to assess this study's two a priori substantive hypotheses regarding the potential association of CTRA gene expression with individual differences in language volume and language structure. Following a statistically significant omnibus test, exploratory follow-up analyses assessed the specific sub-dimension parameters that contributed to the overall domain association with CTRA. Domain associations were tested in an a priori-specified sequence beginning with a null model (predicting similar CTRA expression in all participants), followed by the addition of demographic and behavioral covariates, language volume measures, language structure measures, self-report psychometric measures, and leukocyte subset distributions.

Relative to a null model predicting similar CTRA gene expression across all 143 individuals, addition of an a priori-specified set of demographic and health behavior parameters (age, sex, race, BMI, and smoking) significantly increased predictive power ($F(7, 135) = 3.11$, $p = .0045$; alternative likelihood ratio test in Table S2). Follow-up exploratory analyses of individual sub-dimension parameter estimates (Dataset S2, Model 1) found BMI to be the most significant individual contributor to the overall domain association with gene expression.

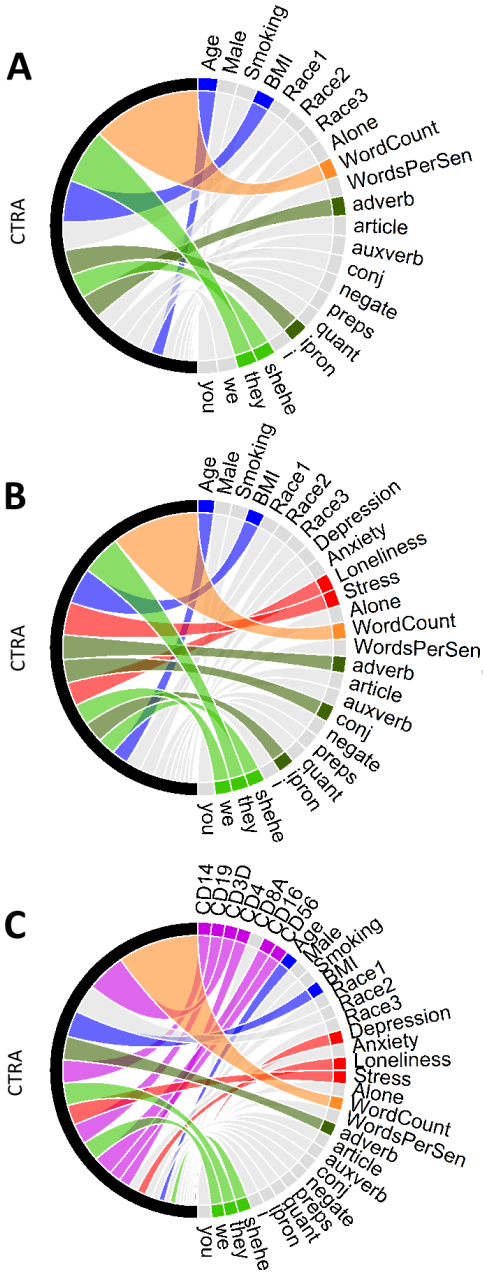

Fig. 2. Prediction of CTRA gene expression. (A) Prediction of peripheral blood mRNA levels for 50 CTRA indicator transcripts (pro-inflammatory genes, interferon response genes, and antibody-related transcripts) by demographic and behavioral characteristics (blue chords), language volume metavariables (orange chords), and language structure features representing the prevalence of 8 categories of function word (dark green chords) and 5 subcategories of personal pronoun (light green chords). Black arcs indicate total CTRA prediction in each mixed effect linear model. Chord widths indicate the relative contribution from standardized values of each variable (squared partial regression coefficients, sorted top-to-bottom by descending magnitude in Dataset S2 Model 3). Colored chords indicate statistically significant effects ($p < .05$) and gray chords indicate non-significant effects. **(B)** Increment to CTRA prediction from adding self-report measures of depression, anxiety, perceived stress, and perceived social isolation / loneliness (red chords); Dataset S2 Model 4. **(C)** Increment to prediction from adding mRNA markers of leukocyte subsets in the peripheral blood cell pool (purple chords); Dataset S2 Model 5.

Addition of basic language volume measures to the model significantly enhanced the prediction of CTRA gene expression beyond the level attainable by demographic and behavioral factors alone ($F(3, 132) = 12.80, p < .0001$). Follow-up exploratory analyses of sub-dimension parameter estimates found low total word count to be the most significant contributor to CTRA prediction by the overall language volume domain (Dataset S2, Model 2).

Inclusion of language structure metrics involving function word prevalence led to a further increase in CTRA prediction beyond that attainable by demographic/behavioral factors and language volume ($F(13, 119) = 6.80, p < .0001$; see also Fig 2a). Follow-up exploratory analyses of individual language structure parameters (Dataset S2, Model 3) found CTRA gene expression to track most strongly a low prevalence of 3rd person plural pronouns (e.g., *they*) and high prevalence of adverbs (e.g., *so*, *very*, *really*), impersonal pronouns (e.g., *it*), and 3rd person singular pronouns (e.g., *she*, *he*).

Ancillary analyses found that language structure metrics also predicted CTRA gene expression in the absence of language volume metrics ($F(13, 122) = 5.12; p < .0001$). Follow-up analyses of individual parameter estimates (Dataset S2, Model 3b) again implicated low prevalence of 3rd person plural pronouns and high prevalence of adverbs, as well as high prevalence of articles*o* and low prevalence of conjunctions (e.g., *and*, *but*).

Psychometric measures

To determine whether language use patterns capture the same predictive information as conventional psychometric instruments, additional analyses included self-report measures of depression, anxiety, perceived stress, and perceived social isolation / loneliness. Results (Fig 2b) continued to show distinct CTRA associations with language volume ($F(3, 112) = 19.24$, $p < .0001$) and language structure ($F(13, 112) = 7.68$, $p < .0001$) above and beyond the significant predictive contribution of psychometric variables ($F(4, 112) = 4.70, p = .0015$). Follow-up exploratory analysis of individual parameters identified loneliness as the only psychometric variable that consistently predicted gene expression (Dataset S2, Model 4). Perceived stress also showed some incremental prediction of gene expression when language metrics were included in the analysis (Dataset S2, Model 4) but not in their absence (Dataset S2, Model 4b). Neither loneliness nor perceived stress showed any significant association with the specific language features that tracked CTRA gene expression (although both were linked to reduced preposition frequency, and loneliness was also associated with a higher prevalence of audio samples in which the participant was alone; Dataset S1).

Leukocyte subsets

CTRA gene expression is structured in part by hematopoietic influences on the distribution of leukocyte subsets in circulating blood (3, 29). As expected, CTRA gene expression varied with an a priori-specified set of 7 variables measuring the prevalence of major leukocyte subset markers ($F(7, 105) = 9.51, p < .0001$; Fig 2c). Follow-up analyses of individual parameters (Dataset S2, Model 5) found this effect to be carried by mRNA markers of monocytes (*CD14*), B lymphocytes (*CD19*), T lymphocytes (*CD3D*, *CD4*), and Natural Killer cells (CD16/*FCGR3A*, CD56/*NCAM1*). However, CTRA gene expression continued to show significant additional associations with language volume ($F(3, 105) = 30.01, p < .0001$) and language structure ($F(13, 105) = 8.67, p < .0001$), as well as psychometric measures ($F(4, 105) = 6.02; p < .0001$), following control for leukocyte subset distributions.

Acute influences

This research was motivated by the hypothesis that stable individual differences in threat-related information processing chronically influence both speech production and gene regulation. However, transient environmental events may also acutely modulate speech and gene expression. To discriminate the relative contributions of acute vs. chronic processes, we compared the strength of association between language patterns and gene

expression in all samples (i.e., both acute and chronic influences) with that observed when acute effects were reduced by excluding blood samples collected within 1 week of language sampling (i.e., chronic influences only). In the absence of acute effects, results continued to show CTRA association with both language volume ($F(3, 102) = 23.04$, $p < .0001$) and language structure ($F(13, 102) = 6.39$, $p < .0001$). In follow-up analyses of the 5 individual language dimensions that significantly predicted gene expression in the overall sample (Dataset S2, Model 3), none showed a significant decrease in predictive strength following the removal of acute influences (Table S3; change in average association strength: +7.4%).

Discussion

This research identified a relationship between individual differences in natural language use and basal gene expression profiles in circulating immune cells. Expression of the CTRA transcriptome profile showed verbal correlates in both the volume of speech (total word count and speech frequency) and the structure of speech (pronoun and adverb prevalence). Patterns of speech production predicted CTRA gene expression substantially better than did self-report measures of negative affective states such as stress, depression, and anxiety, and they did so above and beyond the effects of demographic and behavioral factors (age, sex, race, smoking, BMI) and variations in leukocyte subset distributions. Natural language patterns predicted gene expression even when the two variables were sampled more than 1 week apart, suggesting that their association stems in large part from stable individual differences in processes that jointly influence both speech production and gene regulation. These findings are consistent with the hypothesis that individual differences in automatic CNS threat detection and response systems (6, 30) can influence both language processes (17) and leukocyte gene expression (3, 4). The distinct predictive contributions of self-report psychometric instruments and relatively non-conscious speech patterns parallels the distinct neural substrates of consciously experienced negative affect and the more automatic threat detection and response systems that proximally regulate SNS activity (6) (Fig S1). As such, statistical patterns in natural language use may serve as useful indicators of non-consciously evaluated well-being (i.e., implicit threat vs. safety) that afford greater insight into SNS/β-adrenergic control of peripheral gene expression than do conventional self-report measures of conscious affect.

Exploratory analyses identified several specific language features that predicted CTRA gene expression, but the psychological mechanisms underlying these features remain to be clarified in future research. Low total language output and speech frequency could represent a verbal manifestation of the caution/avoidance and behavioral inhibition responses generated by automatic threat defense responses (6). CTRA down-regulation with increased language output is also consistent with research linking self-expression to physical health (31, 32) and reciprocal data showing effects of psychological inhibition (reduced expression) on SNS activation (33-36). Consistent with that theory, exploratory analyses of other LIWC metrics besides function words identified inhibition-related language as one of the few significant correlates of CTRA gene expression (SI Results: *Exploratory analysis of other LIWC metrics*). Language output may also serve as a proxy for social contact rates that could affect gene expression (13, 37). However, direct measures of the presence vs. absence of others failed to predict CTRA gene expression in this sample, whereas language volume continued to associate with gene expression after control for social contact frequency.

Independent of total language volume, CTRA gene expression also tracked stylistic variations in language structure involving patterns of function word use. We had no a priori hypotheses about the specific types of function word that would associate most strongly with variations in gene expression. Exploratory analyses of that topic found CTRA gene expression to track most strongly a high prevalence of adverbs, impersonal pronouns, and 3rd person singular pronouns, and a low prevalence of 3rd person plural pronouns. Future research will be required to define the psychological mechanisms underlying these specific language features. However, the observed results are consistent with two previous hypotheses regarding the psychological basis for systematic variations in language style. Pronoun use is thought to reflect the deployment of attention (17, 19) and the association of reduced CTRA gene expression with high prevalence of 3rd person plural pronouns (*they*, *them*) may indicate an outward orientation toward the social world that serves to reduce personal threat or arousal and thereby down-regulate SNS activity (38). In contrast, adverbs often serve as intensifiers (*really*, *very*, *certainly*) and their association with up-regulated CTRA gene expression may indicate greater CNS arousal and consequent increases in SNS activity.

Among the psychometric self-report measures examined, only perceived social isolation (loneliness) consistently predicted CTRA gene expression. However, language volume and language structure continued to predict individual differences in CTRA gene expression above and beyond the effects of loneliness and other psychometric characteristics. Consistent with previous research (9, 10, 29, 37), loneliness was associated with up-regulation of the CTRA. In statistical models that included language metrics, perceived stress also showed a counter-intuitive and statistically significant inverse association with the CTRA. Similar inverse stress effects have been observed previously (9, 10) and would be expected in multivariate analyses if gene expression tracks the automatic threat response system whereas perceived stress tracks the functionally distinct but moderately correlated conscious fear/anxiety system (6). Consistent with that prediction, perceived stress showed no significant association with CTRA gene expression in statistical analyses that omitted language metrics.

The present findings are subject to several limitations. The observed relationships involve spontaneous spoken language use sampled from the everyday life of American adults in the Atlanta metropolitan area, and it is unclear whether similar results would emerge in other language contexts (e.g., in writing or topically directed speech) or in other demographic, socioeconomic, or cultural contexts. This is an observational study and additional research will be required to define the causal interactions between language and gene expression. This study was motivated by the hypothesis that speech patterns and leukocyte gene expression profiles come to be associated through their common regulation by stable individual differences in upstream CNS threat detection and response systems (6) (Fig S1). However, language use may also causally affect gene expression (e.g., by influencing perceptual or interpretive processes involved in threat detection or coping responses (6)) and gene expression may causally influence language use (e.g., via effects of circulating cytokines on CNS affective, cognitive, or social processes (39-41) that subsequently impact speech production). This study design was not optimal for detecting the effects of acute environmental influence on language and gene expression due to the broad 2-day language-sampling period, the variable lag between language sampling and blood sampling, and the absence of any experimental manipulation of environmental conditions. The CTRA is specific to immune cells and the language correlates of gene expression in other tissues (e.g., CNS) remain to be determined in future research, as do the health implications of the preset findings. Future studies will also be required to assess the generalizability of the diagnostic language patterns identified here and define their underlying neural and psychological mechanisms (e.g., Do they track threat per se or cognitive coping responses that are activated in response

to threat? Which specific elements of the non-conscious threat defense system interact with language production? Are language patterns a useful proxy measure of intervention-induced changes in SNS activity?). This study focused on a narrow range of a priori-defined language metrics to avoid capitalizing on chance in statistical association analyses. However, language use can be characterized on many other dimensions (17, 42-44) and some of those may also be found to associate with gene expression in future studies (see SI Results *Exploratory analyses of other LIWC metrics* for some initial results). These analyses also tested an a priori genomic hypothesis involving the CTRA gene set as a whole, and no genome-wide discovery analyses were performed to identify specific individual gene transcripts that might relate to language use. Additional genes, gene networks, and epigenetic processes may well be found to relate to language use in future research. These analyses also focused on two a priori hypotheses regarding the general domains of language use that may relate to CTRA gene expression (language volume and language structure), and we had no a priori hypotheses regarding the specific dimensions within each general language domain that might best predict gene expression. Given the exploratory nature of analyses examining specific language features (e.g., word count, adverb prevalence, pronoun frequency), their predictive significance should be regarded as provisional until future studies replicate these findings as a priori hypotheses (45).

The present data indicate a systematic relationship between personal expression and gene expression. As such, statistical pattern analysis of natural language use may provide a useful behavioral indicator of non-consciously evaluated well-being (implicit safety vs. threat) that is both distinct from the information provided by conventional self-report measures and more closely tracks the activity of underlying CNS processes which regulate peripheral physiology, gene expression, and health.

Methods

Study design and data collection. Language, psychometric, and gene expression data were collected under basal conditions from 143 healthy (medication-free) adults in the Atlanta metropolitan area. Speech was sampled over 2 days using an iPod Touch implementation of the EAR (worn on a belt) (23), transcribed by trained coders, and processed through LIWC 2007 software (25) to generate language volume and structure measures. Audio files were coded by raters to exclude spurious data (e.g., asleep) and classify participants as being alone vs. with others (intraclass correlation of 2 independent codings = .94) and talking vs. not talking (.98). Psychometric measures were collected at the time of blood sampling (an average 11 ± SD 9 days after language sampling) and included the Perceived Stress Scale (46) (Cronbach α = .86), Beck Depression Inventory (47) (α = .90), Beck Anxiety Inventory (48) (α = .84), and UCLA Loneliness Scale (49) (α = .91). Total RNA was extracted from resting venous blood samples, tested for mass and integrity, and assayed by Illumina HT-12 v4 BeadArrays in the UCLA Neuroscience Genomics Core Laboratory as previously described (10, 27, 28). Inter-assay CVs for the 50 analyzed gene transcripts averaged 0.78% (range: 0.34%-1.46%). All participants provided written informed consent to participate in the study after the nature and design of the study had been fully described and any related questions had been answered. All procedures were approved by the Institutional Review Board at Emory University.

Analysis. Quantile-normalized gene expression data (GSE87656) were $\log_2$-transformed and standardized within gene for analysis by mixed effect linear models (50) testing association between average expression of the 50 available CTRA indicator transcripts (10, 27, 28) and a priori-specified sets of demographic and health behavior variables (age, sex, race, BMI, smoking), language volume meta-variables (word count, words per sentence, frequency of being alone vs. with others), language structure variables (8 general function word categories and 5 personal pronoun subcategories listed in Dataset S1), psychometric variables (stress, depression, anxiety, loneliness), and 7 RNA transcripts marking the relative prevalence of major leukocyte subsets (*CD14*, *CD19*, *CD3D*, *CD4*, *CD8A*, CD56/*NCAM1*, CD16/*FCGR3A*). Predictors were standardized for comparison of association strength, interferon- and antibody-related transcripts were sign-reversed to reflect their inverse contribution to the CTRA (10, 27, 28), and all models included gene-specific intercepts and a fully parameterized (unstructured) covariance matrix to account for correlation among residuals across the 50 CTRA transcripts. Models were estimated using SAS v9.3 PROC MIXED with omnibus partial *F* tests (26) structured to assess this study's 2 primary substantive hypotheses regarding association of the CTRA gene expression contrast with the 3-dimensional space of language volume measures and the 13-dimensional space of language structure measures (function word prevalence). Parallel likelihood ratio tests (50) were performed to corroborate omnibus *F* tests (Table S2). For predictor domains showing a significant omnibus association, exploratory follow-up tests assessed the specific sub-dimensions contributing to the overall domain association with CTRA. Additional details on measurement and statistical analysis are available in SI Methods.

Acknowledgments

We thank Cindy Chung for her insightful discussions and the UCLA Neuroscience Genomics Core Laboratory for assistance. This research was supported by supported by grants from the National Institutes of Health (R01-AT004698, P30-AG017265, R01-AG043404, R37-AG033590, UL1-TR000454).

1. Cole SW (2013) Social regulation of human gene expression: mechanisms and implications for public health. *Am J Public Health* 103 Suppl 1:S84-92.
2. Cole SW (2014) Human social genomics. *PLoS Genet* 10(8):e1004601.
3. Powell ND, *et al.* (2013) Social stress up-regulates inflammatory gene expression in the leukocyte transcriptome via beta-adrenergic induction of myelopoiesis. *Proc Natl Acad Sci U S A* 110(41):16574-16579.
4. Heidt T, *et al.* (2014) Chronic variable stress activates hematopoietic stem cells. *Nat Med* 20(7):754-758.
5. Beissner F, Meissner K, Bar KJ, & Napadow V (2013) The autonomic brain: an activation likelihood estimation meta-analysis for central processing of autonomic function. *J Neurosci* 33(25):10503-10511.
6. Le Doux J (2015) *Anxious: Using the Brain to Understand and Treat Fear and Anxiety* (Viking, New York).
7. Knight JM, *et al.* (2016) Low Socioeconomic Status, Adverse Gene Expression Profiles, and Clinical Outcomes in Hematopoietic Stem Cell Transplant Recipients. *Clin Cancer Res* 22(1):69-78.
8. Kitayama S, Akutsu S, Uchida Y, & Cole SW (2016) Work, meaning, and gene regulation: Findings from a Japanese information technology firm. *Psychoneuroendocrinology* 72:175-181.
9. Cole SW, *et al.* (2015) Myeloid differentiation architecture of leukocyte transcriptome dynamics in perceived social isolation. *Proc Natl Acad Sci U S A* 112(49):15142-15147.
10. Cole SW, *et al.* (2015) Loneliness, eudaimonia, and the human conserved transcriptional response to adversity. *Psychoneuroendocrinology* 62:11-17.
11. Miller GE, *et al.* (2014) Greater inflammatory activity and blunted glucocorticoid signaling in monocytes of chronically stressed caregivers. *Brain, Behavior, and Immunity*.
12. Antoni MH, *et al.* (2012) Transcriptional modulation of human leukocytes by cognitive-behavioral stress management in women undergoing treatment for breast cancer. *Biological Psychiatry* 71(4):366-372.
13. Creswell JD, *et al.* (2012) Mindfulness-Based Stress Reduction training reduces loneliness and pro-inflammatory gene expression in older adults: A small randomized controlled trial. *Brain Behav Immun* 26(7):1095-1101.
14. Newman ML, Pennebaker JW, Berry DS, & Richards JM (2003) Lying words: predicting deception from linguistic styles. *Pers Soc Psychol Bull* 29(5):665-675.
15. Hancock JT, Curry LE, Goorha S, & Woodworth M (2008) On lying and being lied to: a linguistic analysis of deception in computer-mediated communication. *Discourse Processes* 41(1-23).
16. Skillicorn DB & Little A (2010) Patterns of word use for deception in testimony. *Security Informatics*, eds Yang CC, Chau M, Wang JH, & Chen H (Springer, New York).
17. Pennebaker JW (2011) *The secret life of pronouns: what our words say about us* (Bloomsbury Press, New York).
18. Cohn MA, Mehl MR, & Pennebaker JW (2004) Linguistic markers of psychological change surrounding September 11, 2001. *Psychol Sci* 15(10):687-693.
19. Kacewicz E, Pennebaker JW, Davis M, Jeon M, & Graesser AC (2013) Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology* 33(2):125-143.
20. Pennebaker JW & Lay TC (2002) Language use and personality during crises: Analyses of Mayor Rudolph Giuliani's Press Conferences. *Journal of Research in Personality* 36:271-282.
21. Bock JK (1982) Toward a cognitive psychology of syntax: information proessing contributions to sentence formation. *Psychol Rev* 89(1):1-47.
22. Hickok G (2012) Computational neuroanatomy of speech production. *Nat Rev Neurosci* 13(2):135-145.
23. Mehl MR, Pennebaker JW, Crow DM, Dabbs J, & Price JH (2001) The Electronically Activated Recorder (EAR): a device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers* 33:517-523.
24. Mehl MR, Robbins ML, & Deters FG (2012) Naturalistic observation of health-relevant social processes: the electronically activated recorder methodology in psychosomatics. *Psychosom Med* 74(4):410-417.
25. Pennebaker JW, Booth RJ, & Francis ME (2007) *Linguistic Inquiry and Word Count: LIWC (2007)* (www.liwc.net, Austin, TX).
26. Kutner M, Nachtsheim C, Neter J, & Li W (2005) *Applied Linear Statistical Models* (McGraw-Hill/Irwin, New York) 5 Ed.
27. Fredrickson BL, *et al.* (2013) A functional genomic perspective on human well-being. *Proc Natl Acad Sci U S A* 110(33):13684-13689.
28. Fredrickson BL, *et al.* (2015) Psychological well-being and the human conserved transcriptional response to adversity. *PLoS One* 10(3):e0121839.
29. Cole SW, Hawkley LC, Arevalo JM, & Cacioppo JT (2011) Transcript origin analysis identifies antigen-presenting cells as primary targets of socially regulated gene expression

in leukocytes. *Proc Natl Acad Sci U S A* 108(7):3080-3085.
30. Kagan J (1994) *Galen's prophecy: Temperament in human nature.* (Basic Books, New York).
31. Pennebaker JW (1997) Writing about emotional experiences as a therapeutic process. *Psychological Science* 8(162-166).
32. Gross JJ (2013) *Handbook of Emotion Regulation* (Guilford, New York) 2 Ed.
33. Pennebaker JW & Chew CH (1985) Behavioral inhibition and electrodermal activity during deception. *Journal of Personality and Social Psychology* 49:1427-1433.
34. Gross JJ & Levenson RW (1997) Hiding feelings: the acute effects of inhibiting negative and positive emotion. *J Abnorm Psychol* 106(1):95-103.
35. Gross JJ & Levenson RW (1993) Emotional suppression: physiology, self-report, and expressive behavior. *Journal of Personality and Social Psychology* 64:970-986.
36. Gross JJ (2002) Emotion regulation: affective, cognitive, and social consequences. *Psychophysiology* 39(3):281-291.
37. Cole SW, *et al.* (2007) Social regulation of gene expression in human leukocytes. *Genome Biol.* 8(R189):1-13.
38. Eisenberger NI & Cole SW (2012) Social neuroscience and health: neurophysiological mechanisms linking social ties with physical health. *Nat Neurosci* 15(5):669-674.
39. Dantzer R, O'Connor JC, Freund GG, Johnson RW, & Kelley KW (2008) From inflammation to sickness and depression: when the immune system subjugates the brain. *Nat Rev Neurosci.* 9(1):46-56.
40. Eisenberger NI, Inagaki TK, Mashal NM, & Irwin MR (2010) Inflammation and social experience: an inflammatory challenge induces feelings of social disconnection in addition to depressed mood. *Brain Behav Immun* 24(4):558-563.
41. Inagaki TK, Muscatell KA, Irwin MR, Cole SW, & Eisenberger NI (2012) Inflammation selectively enhances amygdala activity to socially threatening images. *Neuroimage* 59(4):3222-3226.
42. Mehl MR, Vazire S, Holleran SE, & Clark CS (2010) Eavesdropping on happiness: well-being is related to having less small talk and more substantive conversations. *Psychol Sci* 21(4):539-541.
43. Schwartz HA, *et al.* (2013) Personality, gender, and age in the language of social media: the open-vocabulary approach. *PLoS One* 8(9):e73791.
44. Iliev R, Dehghani M, & Sagi E (2015) Automated text analysis in psychology: methods, applications, and future developments. *Language and Cognition* 7(2):265-290.
45. Cao J & Zhang S (2014) Multiple comparison procedures. *JAMA* 312(5):543-544.
46. Cohen S, Kessler RC, & Underwood Gordon L (1995) Perceived Stress Scale. *Measuring stress: A guide for health and social scientists*, (Oxford University Press, New York).
47. Beck A, Steer R, & Brown G (1996) *Beck Depression Inventory - Second Edition manual* (Psychological Corporation, San Antonio TX).
48. Beck AT, Epstein N, Brown G, & Steer RA (1988) An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology* 56(6):893-897.
49. Russell D, Peplau LA, & Cutrona CE (1980) The revised UCLA Loneliness Scale: concurrent and discriminant validity evidence. *J Pers Soc Psychol.* 39(3):472-480.
50. McCulloch CE, Searle SR, & Neuhaus JM (2008) *Generalized, linear, and mixed models* (John Wiley & Sons, Hoboken NJ).
51. Desbordes G, *et al.* (2012) Effects of mindful-attention and compassion meditation training on amygdala response to emotional stimuli in an ordinary, non-meditative state. *Front Hum Neurosci* 6:292.
52. Illumina Inc. (2011) Imputing Human HT-12 Expression BeadChip Data. in *Technical Note: Systems and Software*.
53. Krzywinski M, *et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19(9):1639-1645.