SUPPLEMENTARY MATERIAL: MATERIALS AND METHODS

Tissue-specific RNA-Seq in Human Evoked Inflammation Identifies Blood and Adipose LincRNA Signatures of Cardio-metabolic Diseases

Running title: LincRNA signatures in evoked inflammation

*Yichuan Liu, *Jane F. Ferguson, Chenyi Xue, Rachel L. Ballantyne, Ian M. Silverman, Sager J. Gosai, Jacquelyn Serfecz, Michael P. Morley, Brian D. Gregory, *Mingyao Li, *Muredach P. Reilly

Department of Biostatistics and Epidemiology, Cardiovascular Institute, Department of Biology, Perelman School of Medicine and School of Arts and Science at the University of Pennsylvania, Philadelphia, PA 19104

* These authors contributed equally

METHODS

Study subjects

Each clinical study was performed with approval of the University of Pennsylvania (U.Penn) Institutional Review Board after written informed consent was obtained from all research participants. The Genetics of Evoked-responses to Niacin and Endotoxemia (GENE) study is a single-center U.Penn-based National Institute of Health-sponsored experimental endotoxemia protocol (NIH.gov clinical trial NCT00953667) [1]. Briefly, the GENE study recruited healthy African-American or Caucasian individuals (N = 284, 33% African Americans, age 18-45) to an inpatient protocol that included a pre-LPS acclimatization phase, administration of intravenous LPS bolus (1ng/kg U.S. standard reference endotoxin), and a 30 hour post-LPS phase. Multiple clinical and biochemical variables were recorded. Whole blood RNA samples were collected in Qiagen PAX gene tubes at baseline, 1, 2, 4, 6, 12, and 24 hours post-LPS. Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood immediately after collection using density gradient centrifugation (Ficoll-Paque Plus, GE Healthcare Bio-Sciences, Piscataway, NJ) and PBMC material was stored at -80C until further processed [1, 2]. Samples of gluteal subcutaneous fat tissue were obtained at baseline, 4, 12 and 24 hours following LPS as previously described [3, 4] and snap-frozen for subsequent RNA extraction. In this report, we focus initially on a self-reported European Ancestry (EA) individual (subject A) for deep RNA-seq. To enrich discovery of evoked transcriptome responses, subject A was chosen from those with above median inflammatory response to LPS (**Table 1**). RNA-seq of PBMCs of subject A before and after LPS was performed to compare to whole blood RNA-seq data. Selective findings in subject A were assessed for replication in additional GENE study participants (n=6) chosen also from those with above average inflammatory response **(Table 1)**. The broader generalizabilty of findings in subject A was examined by RNA-seq of additional GENE subjects (n=7) **(Suppl Table 1)**. Based on previous mRNA profiling [3, 4] we selected baseline and 2-hr blood samples, and baseline and 4-hr adipose samples for RNA-seq.

For preliminary clinical translation, we used adipose from lean (healthy volunteers, n=12; mean BMI 24.3) and severely obese patients (bariatric surgery n=12; mean BMI 48.3) as described [4].

RNA-seq library preparation and sequencing

RNA samples of subject A were extracted using RNeasy Lipid Tissue total RNA mini kit (Qiagen, Valencia, CA). Extracted RNA samples underwent quality control (QC) assessment using the Agilent Bioanalyzer (Agilent, Santa Clara, CA) and all RNA samples submitted for sequencing had an RNA Integrity Number (RIN) >6, with a minimum of 1µg input RNA. Poly-A library preparation and sequencing were performed at the Penn Genome Frontiers Institute's High-Throughput Sequencing Facility per standard protocols. Briefly, we generated first-strand cDNA using random hexamer-primed reverse transcription, followed by second-strand cDNA synthesis using RNase H and DNA polymerase, and ligation of sequencing adapters using the TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA). Fragments of ~350 bp were selected by gel electrophoresis, followed by 15 cycles of PCR amplification. The prepared libraries were then sequenced using Illumina's HiSeq 2000 with four lanes per sample which generated 2×101 bp paired-end reads generating ~1,000 million unfiltered reads for adipose and blood samples. We also performed RNA-seq of replicate samples from subject A at lower sequencing depth ("low" depth; ~100 million unfiltered reads/sample) for direct comparison of findings at different RNA-seq depth. RNA-seq data of PBMCs from subject A before and after LPS (~100 million unfiltered reads/sample) were compared to subject A blood RNA-seq data. RNA-seq data (~150-200 million unfiltered reads/sample) of blood of seven independent GENE participants was used to assess generalizability of blood RNA-seq findings in subject A.

Alignment of RNA-seq reads

The RNA-seq data were aligned to the hg19 reference genome using Tophat [5] with default options. We applied several filtering steps; (1) the mapping quality score of each read is ≥ 30, (2) reads from the same pair were mapped to the same chromosome with expected orientations and mapping distance between the read pair was < 500,000 bp, and (3) each read was uniquely mapped. For all blood RNA-seq samples (subject A and others), we eliminated hemoglobin gene sequences because these sequences are uninformative (for LPS modulation) yet dominate the blood RNA-seq data due to their high abundance. All subsequent analyses were based on filtered alignment files. Mapping statistics are summarized in **Suppl Table 2A&B**.

Random sampling of aligned RNA-seq reads

To investigate the effect of sequencing depth on analysis of lincRNAs, we randomly selected reads from the filtered alignment "500M" RNA-seq datasets and created subsets with 5M, 10M, 15M, 20M, 25M, 50M, 75M, 100M, 150M, 200M, 300M, 400M reads for both the pre-LPS and post-LPS blood and adipose samples. Treating the 500M datasets as the gold standard, we evaluated percentages of expressed lincRNAs and LPS-modulated lincRNAs detected at lower sequencing depths.

Detection of change in expression of LPS-modulated linc-RNA and protein coding genes

Transcripts were assembled using Cufflinks[6]. For linc-RNAs, we obtained the locations of 4,662 pre-defined linc-RNAs based on RNA-seq data from 24 different human tissues and that also met the following criteria; at least one isoform of the linc-RNA was reconstructed in at least two different tissues or by two assemblers in the same tissue[7]. For consistency across tissues, we included linc-RNAs that were expressed at either time point at FPKM >0. For each linc-RNA or protein-coding gene, we compared the estimated expression level pre- and post-LPS within the same tissue for a given individual using the cuffdiff option in Cufflinks version 1.3.0. Our analysis of differential expression (i.e., LPS-modulation) involves only one sample in each group. Although a few software packages are available, our own investigations based on simulations (data not shown), indicates that Cufflinks v1.3.0 provides the best balance of power and type I error for our current analyses among all software packages we compared.

Targeted technical validation and external replication

For lincRNA findings in subject A, we also performed selective validation (in subject A) and replication (6 additional GENE subjects, B-G) using qRT-PCR. Primers and probes were designed using the Life Technologies Custom TaqMan Assay Design Tool (**Suppl Table 6)**.

Cell Studies

Quantitative real-time PCR (qRT-PCR) was performed for selective linc-RNAs in primary human adiopocytes and primary human monocytes and macrophages as described in Shah et al. [4] and summarized below.

Primary human adipocyte studies: Fresh human adipose tissue was obtained from abdominal surgery specimens for isolation of adipocyte and stroma-vascular fractions (SVF) as well as culture of human adipocytes [4, 8]. Briefly, the specimen was digested with collagenase (Roche Applied Science, IN), filtered, centrifuged, and cellular layer resuspended in oxidation-fermentation (OF) medium [DMEM/F12 plus penicillin/streptomycin, biotin (4 mg/L), and pantothenate (8 mg/L)] with 20% fetal bovine serum and plated at 30,000 cells/cm^2. Once

confluent, cells were differentiated in serum-free differentiation medium [OF media, insulin (20 nM), hydrocortisone (1 μM), dexamethasone (250 nM), human transferrin (10 mg/L), T3 (0.2 nM), isobutylmethylxanthine (500 μM), and PPARγ agonist GW347845 (2 μM), (gift from GlaxoSmithKline, King of Prussia, PA)] until 80% mature (7-10 days). Cells were washed, and treated ± LPS 100 ng/ml in DMEM/F12 with 0.1% FBS plus penicillin/streptomycin for 2, 4, 12, and 24 hours for three independent experiments, in triplicate. Unless noted, reagents were from Sigma-Aldrich, St Louis, MO.

Primary human macrophage studies: Human monocytes (peripheral blood mononuclear cells, >95% expression of CD14 and HLA-DR), were isolated from donor blood following apheresis and elutriation in the Penn Center for AIDS Research. As described in [4, 8, 9], cells were plated at $3x10^5$ cells/cm^2 and cultured in RPMI with 20% FBS, and penicillin/streptomycin, then supplemented with 100 ng/ml MCSF (Sigma, St Louis, MS) to promote differentiation to macrophages over 7 days. In three independent experiments, each in triplicate, macrophages were washed and treated for 16 hours with: RPMI + 5% FBS alone or 100 ng/ml LPS and 20 ng/ml Interferon Gamma (R&D Systems, Minneapolis, MN) for differentiation to inflammatory (M1) phenotype for differentiation to the M1 phenotype, or 20 ng/mL recombinant human IL4 (R&D Systems, Minneapolis, MN) for differentiation to the M2 phenotype [10].

Pilot case-control study of adipose lincRNA expression in obesity

We used adipose from lean (healthy volunteers, n=12; mean BMI 24.3) and severely obese patients (bariatric surgery n=12; mean BMI 48.3) as described described [4]. qRT-PCR was performed for selective linc-RNAs and B-actin mRNA as reference gene. Mann Whitney U non-parametric tests were used to test for differences in the delta CTs for candidate lincRNAs in lean vs. obese adipose.

Analysis of conservation and epigenetic promoter marks

To assess the sequence conservation of lincRNAs, we used the pre-computed phastCons scores [11] from 46 primate genomes (available in conservation track from UCSC genome browser http://genome.ucsc.edu/). For each lincRNA, a set of "control exons" were selected by matching the length of lincRNA exons with intergenic regions on the same chromosome. Additionally, we required the "control exons" to have the same relative positions as exons in the lincRNA. Coding exons were selected from protein coding genes whose numbers of exons match those expressed lincRNAs. Cumulative frequency of phastCons scores were plotted for all expressed lincRNAs expressed in adipose and blood. To assess lincRNA epigenetic features consistent with promoter regulation, we used the ChIP-seq data of H3K4me3 (GM12878 LCL cell line, ENCODE project [12]) to stratify our lincRNAs into those with evidence of active promoter and those without. LincRNAs with H3K4me3 peak within 2.5kb upstream of the first exons were considered to have active promoter.

Analysis of transcription factor binding sites

We hypothesized that LPS-lincRNAs would have more binding sites for transcription factors in the NF-κB family. To test this hypothesis, we obtained the motifs for NFKB1, NFKB2, RELA, RELB and REL from TRANSFAC database (version 2012.3; https://portal.biobase-international.com/build_t/idb/1.0/html/bkldoc/source/bkl/transfac%20suite/transfac/tf_citing.html)[13]. Sequences encompassing 2.5kb upstream and 300bp downstream of a lincRNA's start site were searched for motif occurrences using FIMO [14]. Fisher's exact test was used to test for enrichment of motifs for each transcription factor. We also calculated the number of motifs in

the 2.8kb region for each lincRNA that has at least 1 motif. Two-sample t-test was conducted for LPS-modulated and non LPS-modulated lincRNAs in blood and adipose.

Analysis of GWAS summary data for cardio-metabolic traits

We interrogated GWAS signals summarized in the NHGRI GWAS catalogue (http://www.genome.gov/gwastudies; accessed June 28 2013) for overlap with lincRNAs that were LPS-modulated in blood or adipose. We focused on selective GWAS signals for inflammatory and cardio-metabolic traits (**Suppl Table 5**).

REFERENCES

1. Ferguson JF, Patel PN, Shah RY, Mulvey CK, Gadi R, Nijjar PS, Usman HM, Mehta NN, Shah R, Master SR, Propert KJ, Reilly MP. Race and gender variation in response to evoked inflammation. *J Transl Med*. 2013;11:63
2. Anderson PD, Mehta NN, Wolfe ML, Hinkle CC, Pruscino L, Comiskey LL, Tabita-Martinez J, Sellers KF, Rickels MR, Ahima RS, Reilly MP. Innate immunity modulates adipokines in humans. *J Clin Endocrinol Metab*. 2007;92:2272-2279
3. Mehta NN, McGillicuddy FC, Anderson PD, Hinkle CC, Shah R, Pruscino L, Tabita-Martinez J, Sellers KF, Rickels MR, Reilly MP. Experimental endotoxemia induces adipose inflammation and insulin resistance in humans. *Diabetes*. 2010;59:172-181
4. Shah R, Lu Y, Hinkle CC, McGillicuddy FC, Kim R, Hannenhalli S, Cappola TP, Heffron S, Wang X, Mehta NN, Putt M, Reilly MP. Gene profiling of human adipose tissue during evoked inflammation in vivo. *Diabetes*. 2009;58:2211-2219
5. Trapnell C, Pachter L, Salzberg SL. Tophat: Discovering splice junctions with rna-seq. *Bioinformatics (Oxford, England)*. 2009;25:1105-1111
6. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28:511-515
7. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding rnas reveals global properties and specific subclasses. *Genes & development*. 2011;25:1915-1927
8. Shah R, Hinkle CC, Ferguson JF, Mehta NN, Li M, Qu L, Lu Y, Putt ME, Ahima RS, Reilly MP. Fractalkine is a novel human adipochemokine associated with type 2 diabetes. *Diabetes*. 2011;60:1512-1518
9. Lehrke M, Reilly MP, Millington SC, Iqbal N, Rader DJ, Lazar MA. An inflammatory cascade leading to hyperresistinemia in humans. *PLoS Med*. 2004;1:e45
10. Martinez FO, Sica A, Mantovani A, Locati M. Macrophage activation and polarization. *Front Biosci*. 2008;13:453-461
11. Siepel A, Bejerano G, Pedersen JS, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005;15:1034-1050
12. Consortium EP, Dunham I, Kundaje A, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57-74
13. Matys V, Kel-Margoulis OV, Fricke E, et al. Transfac and its module transcompel: Transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*. 2006;34:D108-110
14. Grant CE, Bailey TL, Noble WS. Fimo: Scanning for occurrences of a given motif. *Bioinformatics (Oxford, England)*. 2011;27:1017-1018