



Published in final edited form as:

Shock. 2018 July ; 50(1): 53–59. doi:10.1097/SHK.0000000000001029.

GRAPH THEORETICAL ANALYSIS OF GENOME-SCALE DATA: EXAMINATION OF GENE ACTIVATION OCCURRING IN THE SETTING OF COMMUNITY-ACQUIRED PNEUMONIA

Octavia M. Peck Palmer^{1,2,3,4}, Gary Rogers⁵, Sachin Yende^{2,3,4}, Derek C. Angus^{2,3,4},
Michael A. Langston⁵, and Gilles Clermont^{2,3}

¹Department of Pathology, University of Pittsburgh, Pittsburgh, PA, 15213, USA

²Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, 15213, USA

³The Clinical Research, Investigation, and Systems Modeling of Acute Illness (CRISMA) Center, Pittsburgh, PA, 15213, USA

⁴Department of Clinical and Translational Science, Pittsburgh, PA, 15213, USA

⁵Department of Electrical Engineering & Computer Science, University of Tennessee, Knoxville, TN, 37996, USA

Abstract

Introduction—We have previously reported evidence that black individuals appear to have a significantly higher incidence of infection-related hospitalizations compared to white individuals. It is possible that the host immune response is responsible for this vital difference. In support such a hypothesis, the aim of this study was to determine whether black and white individuals exhibit differential whole blood gene network activation.

Methods—We examined whole blood network activation in a subset of patients (n=22 pairs, propensity score matched (1:1) black and white patients) with community-acquired pneumonia (CAP) from the Genetic and Inflammatory Markers of Sepsis study. We employed day 1 whole blood transcriptomic data generated from this cohort and constructed co-expression graphs for each racial group. Pearson correlation coefficients were used to weight edges. Spectral thresholding was applied to ascribe significance. Innovative graph theoretical methods were then invoked to detect densely connected gene networks and provide differential structural analysis.

Results—Propensity matching was employed to reduce potential bias due to confounding variables. Although black and white patients had similar socio- and clinical demographics, we identified novel differences in molecular network activation – dense subgraphs known as paracliques that displayed complete gene connection for both white (three paracliques) and black patients (one paraclique). Specifically, the genes that comprised the paracliques in the white patients include circadian loop, cell adhesion, mobility, proliferation, tumor suppression, NF κ B, and chemokine signaling. However, the genes that comprised the paracliques in the black patients

include DNA and mRNA processes, and apoptosis signaling. We investigated the distribution of black paracliques across white paracliques. Black patients had five paracliques (with almost complete connection) comprised of genes that are critical for host immune response widely distributed across 22 paracliques in the white population. Anchoring the analysis on two critical inflammatory mediators, IL-6 and IL-10 identified further differential network activation among the white and black patient populations.

Conclusions—These results demonstrate that, at the molecular level, black and white individuals may experience different activation patterns with CAP. Further validation of the gene networks we have identified may help pinpoint genetic factors that increase host susceptibility to community-acquired pneumonia, and may lay the groundwork for personalized management of CAP.

Keywords

race; graph theoretical; community acquired pneumonia; inflammatory; microarray

INTRODUCTION

Globally, community-acquired pneumonia (CAP), a common infectious disease of the respiratory tract, accounts for significant morbidity and mortality (1, 2). In the United States, black individuals have a higher incidence of CAP that requires hospitalization and are younger at disease onset than their white counterparts (1, 2). Overall, adjustment for crude social and economic factors has not fully explained the CAP disparities and suggests an underlining biological basis may be at play (3–6). Racial differences in the epidemiology of CAP may, in part, be a functional consequence of black and white populations mounting a different immune response to infection. However, identification of molecular mechanisms that modulate racial differences in the incidence and outcomes related to CAP will enhance our understanding of CAP and may lay the groundwork for personalized management of CAP.

Previous studies that examined racial differences in the immune response to infection focused mainly on inflammatory biomarker concentration differences and the identification of single nucleotide polymorphisms (SNPs) within candidate genes (7, 8). Few studies of CAP have examined SNPs within candidate genes with susceptibility, illness severity, and outcomes among different racial groups (9). CAP is of course complex and so studying genes in isolation may not clearly delineate connections between genes. Thus, statistical interrogation of co-expressed genes and molecular gene network activation may help identify key factors that mediate differences in host immune response to infection among racial populations.

Previously, our laboratory developed graph theoretical algorithms for the study of problems such as this. Of central interest is the generation of paracliques, which are clique-centric, noise resistant, dense subgraphs (10, 11). We have applied these novel methods in a variety of settings to identify sets of co-expressed genes (10, 12–14). For example, in Voy et al. graph theoretical algorithms were employed and used to demonstrate in a murine model of *in vivo* lose radiation exposure differential molecular pathway activation (13). The present

study is the first to apply graph theoretical approaches to interrogate host immune response in CAP. We analyzed a set of transcriptomic data (whole blood samples) that generated from propensity score matched (one black patient matched with one white patient with similar score) black and white patients with CAP admitted to the emergency department (ED). The aims of the study were to determine if i) Black and white patients with CAP that have similar socio- and clinical baseline demographics exhibit different co-expressed genes at ED admission, ii) The distribution of co-expressed genes among the populations differ and iii) IL-6 and IL-10 associated co-expressed genes were different among the populations.

MATERIALS AND METHODS

Patient Selection and Propensity Score Model

We previously described the clinical design, patient demographics, and clinical definitions of the GenIMS study (15). Briefly, the GenIMS study (from 2003–2006) is an inception cohort study of 2,320 patients enrolled thru the ED (28 US academic and community hospitals) with clinical and radiologic diagnoses of CAP (15). Our study cohort was comprised of 44 patients (n=22 black and n=22 white patients). The patients self-reported their race. The patients were selected and matched based on propensity scores and represented 12 of the 28 enrollment sites. All 12 enrollment sites were within western Pennsylvania. Racial differences in immune host response may be confounded by differences in pre-infection factors, such as age, sex, health behaviors, and comorbidities (6). To minimize this confounding, we used logistic regression and constructed a “score” for the propensity to be black in the GenIMS study based on three pre-infection patient characteristics: age, sex, and pack of cigarettes smoked per day (6). In our study cohort, we used the propensity scores to perform a 1:1 match of black and white patients.

Whole blood mRNA analysis for transcriptomic profile identification

Patients had whole blood samples collected at ED admission. Each patient sample was analyzed in duplicate or triplicate using the Human RefSeq8 Expression BeadChips. The BeadChips targeted 24,500 well-annotated transcripts per patient sample (Illumina BeadChip Technology platform). Total mRNA was extracted from the whole blood sample using the Roche mRNA isolation kit for Blood/Bone Marrow (Roche, Indianapolis, IN). Quantification for mRNA via Ribogreen (Life Technologies, Santa Clara, CA) analysis was performed to eliminate DNA carry over (contamination), and mRNA integrity was assessed. We prepared samples for Illumina BeadChip® analysis using the Illumina RNA amplification kit available from Ambion. First strand cDNA was prepared from 200 ng total RNA using T7 Oligo(dT) primers, dNTP mix, first strand buffer and ArrayScript enzyme. We incubated the reaction at 42°C for 2 hours. Second strand cDNA synthesis was performed by adding second strand buffer, extra dNTPs, RNase H and DNA polymerase I and incubating for a further 2 hours at 16°C. Double stranded cDNA was cleaned per manufacturer’s instructions. Briefly, the cDNA binding buffer was added to the reaction, and this mixture was applied to the cleanup column. Following centrifugation, the column was washed once with DNA wash buffer and excess wash buffer removed via centrifugation. For maximum yield columns were eluted twice with equal volumes of warm nuclease free water. Samples were dried in a speed vacuum and *in vitro* transcription master mix (1X reaction

buffer, 7.5 mM each ATP, CTP, GTP, 3.75 mM UTP, 3.75 mM biotin-16-UTP and 1X T7 enzyme) was added. We incubated the reaction for 12 hours at 37°C. Following incubation RNA binding buffer and ethanol were added to the samples which were applied to the cleanup column. The samples were washed again and eluted with 100 ul warm nuclease free water. The samples were quantified using spectrophotometry and concentrated in a speed vacuum. The appropriate mass of cRNA was mixed with EB1 hybridization buffer (supplied by the manufacturer. Hybridization chambers for each BeadChip were assembled as per manufacturer's instructions. Sample mixture was applied to each chamber. BeadChips were hybridized overnight at 55°C with rotation. Following hybridization, the BeadChips were washed with manufacturer supplied wash buffer. Blocking was performed at room temperature using supplied blocking solution, and the BeadChips were stained with streptavidin –Cy3 for 10 minutes. Excess stain solution was washed off with provided wash buffer, and Beadchips were dried via centrifugation. Dry BeadChips were scanned using the BeadArray Reader and analysis performed with Illumina Bead Studio application. Expression array raw data was analyzed using RMAExpress, and Biometric Research Branch ArrayTools were used to perform the following on the raw data: quality assessment, probe analysis, and normalization. DAVID, an online bioinformatics tool was used to determine gene ontology. The microarray data for this study adhered to the Minimal Information About a Microarray Experiment (MIAME) guidelines (23). The raw and normalized data are available from the corresponding author upon request.

Graph theoretical analysis

Differential correlations (12, 13) were computed. The upper edge weight cutoff was set at 0.90 and the lower at 0.30. In this fashion, 7,898 gene pairs (edges) of possible interest were discovered (10, 14). We then computed all maximal cliques in the resultant differential graph and divided the transcriptomic data into two racial patient subsets. Paracliques (10, 14) were generated for black and white patient groups separately, at thresholds 0.91 and 0.89, respectively, using spectral methods (16). To determine the specificity of the paracliques to each racial group, we examined the distribution of the black patient population paracliques across the white patient population paracliques.

Assessment of noise in paraclique analysis

Due to the modest size of the study population, we sought to describe better the level of noise associated with paraclique identification and comparison. We compared results contrasting white and black paracliques to results that contrasting paracliques obtained from random subgroups of patients. Specifically, we created 50 pairs of cohorts, where each cohort consisted of 22 randomly selected (without replacement) patients, eleven white and eleven black patients. We chose to balance black and white patients within cohorts of a pair to avoid a systematic race-related effect to bias results. Ideally, paracliques computed for cohorts of a pair would overlap significantly. We constructed nine metrics to compute paraclique overlap within pairs, computed paracliques for all pairs, and therefore created a distribution for each overlap metric across the 50 pairs. Finally, the same metric was computed for paracliques from black and white patients at the same correlation threshold level. We computed the z-score of each overlap metric from the distribution of random pairs

and a p-value generated for each metric to assess the significance of difference observed between black and white paracliques.

Genetic Estimation of African and European Ancestry

Since black Americans in the US are an admixed population, meaning they are of both African and European ancestry, we genotyped all the black patients in our study cohort for 93 ancestry informative markers as previously described (17). These ancestry informative markers allowed us to genetically estimate the proportion of African and European ancestry for each black patient using a maximum likelihood estimation procedure. Quantification of African and European ancestry for the black patients was used to help us understand how ancestry contributes to racial differences seen in immune response to infection.

RESULTS

Black and White patients exhibit similar clinical characteristics at ED admission but different inflammatory mediator concentrations

The propensity-matched study cohort showed no significant differences in age (mean ages, 52.5 years old, black patients, vs. 50.1 years old, white patients, $p=0.63$, Table 1) or symptom duration before ED admission (6.4 days, black patients vs. 5.8 days, white patient, $p=0.62$, Table 1). Furthermore, no differences were seen in illness severity at ED admission (Charlson score =2.09, $p=0.21$; ED admission Acute Physiology And Chronic Health Evaluation III score, 35 ± 14 , black patients, vs. 35 ± 8 , white patients, $p=0.43$, ED admission Table 1). However, Black patients ($n=22$) had lower circulating IL-6 (29.6 ± 2.3 vs. 43.3 ± 2.1 pg/ml) and IL-10 (3.2 ± 0.2 vs. 5.4 ± 0.5 pg/ml) concentrations at ED admission compared to white patients ($n=22$), respectively. While WBC and granulocytes tended to be higher in White patients (15.3 ± 12.0 vs. $10.3 \pm 5.5 \times 10^9/l$ and 12.7 ± 11.9 vs. $8.2 \pm 5.6 \times 10^9/l$), only monocyte counts were statistically higher ($.69 + .32$ vs. $.41 + .32 \times 10^9/l$). The self-identified black patients had genetic ancestry estimates of 73.8% African and 26.1% European whereas the self-identified white patients had genetic ancestry estimates of 91.1% European and 8.2% African.

Identification of gene networks with complete connection in a single population

We identified four paracliques that were population specific – meaning that one population expressed the corresponding molecular network and the other did not. In the black population, we identified one ‘complete connection’ paraclique comprised of 16 genes not shared in networks in the white population (Table 2). The genes comprising the aforementioned paraclique are involved in DNA replication/repair, mRNA elongation, mitosis/meiosis, and apoptosis signaling. We identified the white population three ‘complete connection’ paracliques (comprised of 10, 11, 14 genes) (Tables 3, 4, and 5 respectively) comprised of genes involved in the circadian loop, cell adhesion, mobility, proliferation, tumor suppression, phosphate metabolism, cholesterol efflux, smooth muscle cell proliferation, nuclear factor kappa beta (NF κ B) signaling, chemokine signaling, vasculogenesis and angiogenesis.

Paracliques identified in the black population are widely distributed among paracliques identified in the white population

In addition to identifying paracliques with complete connection specific to the black and white patient groups, we further identified paracliques that were comprised of genes involved in the host immune response and implicated in the pathogenesis of CAP. Specifically, paraclique #1 had a density of 0.999074, contained 81 genes that were involved in cell cycle and regulation signaling, metallothionein 1E binding was spread across 10 white paracliques (Figure 1. a). Paraclique #2 had a density of 0.989247, was comprised of 31 genes that were involved in apoptosis, Interleukin -17D, leptin, iodide transporter signaling was spread across 4 white paracliques (Figure 1. b). Paraclique #3 had a density of 0.991342 and comprised of 22 genes involved in lymphocyte development, maturation, and adhesion spread across 5 white paracliques (Figure 1. c). Paraclique #233 had a density of 1 and comprised of 3 genes involved in lipid raft signaling and ferritin and metal ion transport was spread across 2 white paracliques (Figure 1. d). Paraclique #237 had a density of 1 and was comprised of 3 genes involved in apoptosis, and acute monocyte/neutrophil inflammatory responses spread across 1 white paracliques (Figure 1. e).

The distribution of the five Black paracliques comprised of genes involved in apoptosis, cell cycle regulation, lymphocyte development, and metal transport pathways across white paracliques was investigated. The five black paracliques distributed across 22 white paracliques (Figure 1. a–e) suggesting the black paracliques are highly distributed across white paracliques.

Gene networks identified in the black population are distinct paracliques

To further confirm that the paracliques identified in the black population were distinct, we selected two paracliques, paraclique #3 and #237 in the black population for two primary reasons. First, the fore-mentioned paracliques contain genes that are broadly distributed among several paracliques identified in the white population data. This broad distribution suggests a divergence in network signaling across the two populations. Secondly, these paracliques are comprised of genes that have been implicated in infection and sepsis. We calculated the corresponding correlation coefficients. Next, we determined the threshold at which the two paracliques coalesced. Our results showed that the correlation value had to be lowered to 0.0014 before the black paracliques coalesced. Thus the two paracliques in the black population were distinct paracliques (data not shown).

Noise characterization

Nine metrics were computed to quantify the overlap between black and white paracliques (Supplementary Table 1). Of those, four were significantly different from the mean overlap computed from 50 random pairs; three more had $0.05 < p < 0.10$ and the last two pairs had $0.10 < p < 0.2$. Note that paraclique computations are designed to handle noise, and thus white and black paracliques showed significantly less overlap than paracliques from randomly chosen pairs of cohorts. Modifying the correlation threshold to 0.95 did not improve these results.

Gene networks highly correlated with IL-6 and IL-10

We further characterized the gene networks within each patient group. IL-6 and IL-10 are critical inflammatory mediators of a robust immune response. In the present study, we identified significantly lower circulating IL-6, and IL-10 in black patients at ED admission compared to white patients and this was also exhibited in the original larger cohort (6). Thus, we examined the gene networks specific to IL-6 and IL-10 signaling. We anchored the analysis on IL-6 and IL-10 (Supplementary Table 2 a,b $p < 0.01$) and identified distinct network activation. Partial correlations were computed between IL-6 and all other genes, controlling for IL-10, and between IL-10 and all other genes, controlling for IL-6. Genes with correlation p -values at or below 0.01 were selected. Black and white patients did not share common gene networks. Black patients exhibited a genetic network comprised of 16 genes involved in protein coding, transcription, and RNA transport/splicing (Supplementary Table 2). On the other hand, white patients showed a genetic network comprised of 26 genes, involved in protein coding, IL-17A, calcium transport, and apoptosis signaling (Supplementary Table 2b).

DISCUSSION

The present study, to our knowledge, is the first study to use graph theoretical analysis of genome-scale data to examine whole blood gene network activation among black and white patients with CAP. In our cohort of CAP, black and white patients with similar clinical characteristics showed differential network activation at ED admission. We identified paracliques comprised of genes in ‘complete connection’ – no genes shared (connectionless) for each population. The white patients exhibited three paracliques and the black population exhibited one paraclique with ‘complete’ connection. Our data is one such example of differential topology (12), a putative gene response network that is completely activated in one patient population but not activated in the other population.

Examining of gene network activation is important in understanding etiology and disease states. Mootha et al. demonstrated the clinical importance of identifying co-expressed genes by showing that the skeletal muscle gene network activation was significantly different between healthy volunteers compared to individuals with diabetes and was associated with insulin, oxidative and metabolic (18). Wolen et al. used paracliques to identify ethanol-activated gene networks in a murine model – novel genotype/phenotype associations illustrated by DNA changes and ethanol sensitivity (19). In the present study, we showed that black patients exhibited five novel molecular networks comprised of genes involved in innate/adaptive immunity and the genes distributed across 22 novel networks specific to the white patients. This finding suggests that in the setting of CAP the gene network signal identified in the black population is decomposed across several gene networks in the white population. Our laboratory and others have identified distinct paracliques in different disease settings (20, 21), including Type 1 diabetes (22), seasonal allergic rhinitis (21), and cancer (23).

Our study is built on several major strengths. We collected whole blood at ED admission to examine early host immune differences before the patient receiving antibiotic treatment. We examined racial differences in host immune response using a cohort of black patients; a

population often underrepresented in the biomedical studies. We also included host immune response differences using CAP as the sole infectious process to minimize the confounding effect of different sources of infections. We employed robust statistical approaches that included propensity matching to minimize the confounding effects of non-genetic factors and applied graph theoretical methods (14). Another strength of our approach was the extraction of paracliques, which are noise-resistant subgraphs containing as many edges as possible (10, 14). Density thus greatly reduces opportunities for false positives, and among current methods is known to provide the greatest biological fidelity using well-annotated data such as that produced from *S. cerevisiae* (11).

Our study also has limitations. Due to logistical reasons we were only able to examine day 1 transcriptomic changes in the present study. Transcriptomic changes throughout the course of hospitalization for CAP are unknown. We were unable to determine the specific pathogen types and pathogen load for each patient in our cohort. The transcriptomic results were generated from whole blood specimens and represent a mixed cell population. Populations of African descent reportedly exhibit lower neutrophil counts compared to European populations (22). It is plausible that differences in specific cell populations may contribute to some of the molecular network differences exhibited between populations, as we also observe some differences in differential counts between White and Black patients. Thus, our molecular network findings may provide insight into which cell populations are of importance for further analysis. Our results suggest that future studies should pay particular attention to circulating monocytes. Fundamental differences may also exist in genetic programming between black and white patients and may partly explain the racial differences we observe beyond CAP. The addition of a control group of healthy individuals may identify this possible confounding factor.

We are aware that 22 patients per cohort is a relatively small sample size. To help counter the noise we employed noise-resistant procedures such as the paraclique algorithm. Moreover, the process of creating random pairs of cohorts undid, to a certain extent, the careful matching we applied in generating the initial black and white cohorts. Therefore, we suspect we might, in fact, have overestimated noise although this is difficult to quantify more precisely. Nevertheless, black and white paracliques were significantly more disjointed than random pairs, further supporting the main conclusions of our analysis.

In conclusion, our data suggest that in the setting of CAP, differential gene activation occurs among the patients. These findings warrant further validation in a larger cohort of community-acquired pneumonia focused on cell-type specific gene activation which may help delineate genetic factors that increase host susceptibility to CAP and long-term outcomes, and may lay the groundwork for personalized management of CAP. An increased understanding of these networks and any differences across populations has the potential to help target differential therapies for both treatment and prevention of CAP, dependent of course on how such therapies evolve.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding source: OPP is supported by KL2 RR02415 and MAL is supported by R01AA018776.

We wish to express our appreciation to the GenIMS investigators and patient participants, Arnold M. Saxton, PhD, for his help with statistical analysis, Chelsea A. Wells, MD, for her assistance with genome research and table development, and Joanne H. Hasskamp, MS, for her assistance with manuscript formatting and reference sourcing.

References

1. Kaplan V, Angus DC, Griffin MF, Clermont G, Scott Watson R, Linde-Zwirble WT. Hospitalized community-acquired pneumonia in the elderly. *Am J Respir Crit Care Med.* 2002; 165(6):766–772. [PubMed: 11897642]
2. Wiemken TL, Peyrani P, Ramirez JA. Global changes in the epidemiology of community-acquired pneumonia. *Semin Respir Crit Care Med.* 2012; 33(03):213–219. [PubMed: 22718207]
3. Marston BJ, Plouffe JF, File TM Jr, Hackman BA, Salstrom SJ, Lipman HB, Kolczak MS, Breiman RF, for the Community-Based Pneumonia Incidence Study Group. Incidence of community-acquired pneumonia requiring hospitalization: results of a population-based active surveillance study in Ohio. *Arch Intern Med.* 1997; 157(15):1709–1718. [PubMed: 9250232]
4. Mayr FB, Yende S, Linde-Zwirble WT, Peck-Palmer OM, Barnato AE, Weissfeld LA, Angus DC. Infection rate and acute organ dysfunction risk as explanations for racial differences in severe sepsis. *JAMA.* 2010; 303(24):2495–2503. [PubMed: 20571016]
5. Jha AK, Stone R, Lave J, Chen H, Klusaritz H, Volpp K. The concentration of hospital care for black veterans in Veterans Affairs hospitals: implications for clinical outcomes. *Journal for Healthcare Quality.* 2010; 32(6):52–61. [PubMed: 20946426]
6. Palmer OM, Yende S, Song C, Mayr F, Weissfeld L, Angus D, Tseng G. Whole blood transcriptomic signatures differ between ethnic populations with community acquired pneumonia and severe sepsis. *Shock.* 2012; 37(Supplement 1):1. [PubMed: 22157941]
7. Mayr FB, Spiel AO, Leitner JM, Firbas C, Jilma-Stohlawetz P, Chang JY, Key NS, Jilma B. Racial differences in endotoxin-induced tissue factor-triggered coagulation. *J Thromb Haemost.* 2009; 7(4):634–640. [PubMed: 19187081]
8. Mayr FB, Spiel AO, Leitner JM, Firbas C, Kliegel T, Jilma-Stohlawetz P, Petra MD, Derendorf H, Jilma B. Duffy antigen modifies the chemokine response in human endotoxemia. *Crit Care Med.* 2008; 36(1):159–165. [PubMed: 18007272]
9. Yende S, Angus DC, Kong L, Kellum JA, Weissfeld L, Ferrell R, Finegold D, Carter M, Leng L, Peng Z, Bucala R. The influence of macrophage migration inhibitory factor gene polymorphisms on outcome from community-acquired pneumonia. *FASEB J.* 2009; 23(8):2403–2411. [PubMed: 19346297]
10. Chesler EJ, Langston MA. Combinatorial genetic regulatory network analysis tools for high throughput transcriptomic data. In: Eskin E, Ideker T, Raphael B, Workman C, editors *Systems Biology and Regulatory Genomics*. Berlin, Germany: Springer-Verlag; 2006. 150–165.
11. Langston MA, Levine RS, Kilbourne BJ, Rogers GL, Kershenbaum AD, Baktash SH, Coughlin SS, Saxton AM, Agboto VK, Hood DB, Litchveld MY, Oyana TJ, Matthews-Juarez P, Juarez PD. Scalable combinatorial tools for health disparities research. *Int J Environ Res Public Health.* 2014; 11(10):10419–10443. [PubMed: 25310540]
12. Langston MA, Perkins AD, Saxton AM, Scharff JA, Voy BH. Innovative computational methods for transcriptomic data analysis: a case study in the use of FPT for practical algorithm design and implementation. *The Computer Journal.* 2008; 51(1):26–38.
13. Voy BH, Scharff JA, Perkins AD, Saxton AM, Borate B, Chesler EJ, Branstetter LK, Langston MA. Extracting gene networks for low dose radiation using graph theoretical algorithms. *PLoS Computational Biology.* 2006; 2(7):e89. [PubMed: 16854212]
14. Hagan RD, Langston MA, Wang K. Lower bounds on paraclique density. *Discrete Applied Mathematics.* 2016; 204:208–212. [PubMed: 27057077]
15. Yende S, van der Poll T, Lee M, Huang DT, Newman AB, Kong L, Kellum JA, Harris TB, Bauer D, Satterfield S, Angus DC. The influence of pre-existing diabetes mellitus on the host immune

- response and outcome of pneumonia: analysis of two multicentre cohort studies. *Thorax*. 2010; 65(10):870–877. [PubMed: 20861291]
16. Perkins AD, Langston MA. Threshold selection in gene co-expression networks using spectral graph theory techniques. *BMC Bioinformatics*. 2009; 10(Suppl 11):S4.
 17. Halder I, Shriver M, Thomas M, Fernandez JR, Frudakis T. A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Hum Mutat*. 2008; 29(5):648–658. [PubMed: 18286470]
 18. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003; 34(3):267–273. [PubMed: 12808457]
 19. Wolen AR, Phillips CA, Langston MA, Putman AH, Vorster PJ, Bruce NA, York TP, Williams RW, Miles MF. Genetic dissection of acute ethanol responsive gene networks in prefrontal cortex: functional and mechanistic implications. *PLoS One*. 2012; 7(4):e33575. [PubMed: 22511924]
 20. Schoenrock A, Samanfar B, Pitre S, Hooshyar M, Jin K, Phillips CA, Wang H, Phanse S, Omid K, Gui Y, Alamgir M, Wong A, Barrenäs F, Babu M, Benson M, Langston MA, Green JR, Dehne F, Golshani A. Efficient prediction of human protein-protein interactions at a global scale. *BMC Bioinformatics*. 2014; 15:383. [PubMed: 25492630]
 21. Barrenas F, Chavali S, Alves AC, Coin L, Jarvelin MR, Jornsten R, Langston MA, Ramasamy A, Rogers G, Wang H, Benson M. Highly interconnected genes in disease-specific networks are enriched for disease-associated polymorphisms. *Genome Biology*. 2012; 13(6):R46. [PubMed: 22703998]
 22. Eblen JDGI, Saxton AM, Wu J, Snoddy JR, Langston MA. Graph algorithms for integrated biological analysis, with applications to type 1 diabetes data. In: Butenko S, Chaovalitwongse WA, Pardalos PM, editors *Clustering Challenges in Biological Networks*. Singapore: World Scientific; 2009. 207–222.
 23. Pradhan MP, Prasad NK, Palakal MJ. A systems biology approach to the global analysis of transcription factors in colorectal cancer. *BMC Cancer*. 2012; 12:331. [PubMed: 22852817]
 24. Reich D, Nalls MA, Kao WHL, Akylbekova EL, Tandon A, Patterson N, Mullikin J, Hsueh W, Cheng C, Coresh J, Boerwinkle E, Li M, Waliszewska A, Neubauer J, Li R, Leak TS, Ekunwe L, Files JC, Hardy CL, Zmuda JM, Taylor HA, Ziv E, Harris TB, Wilson JG. Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. *PLoS Genetics*. 2009; 5(1):e1000360. [PubMed: 19180233]

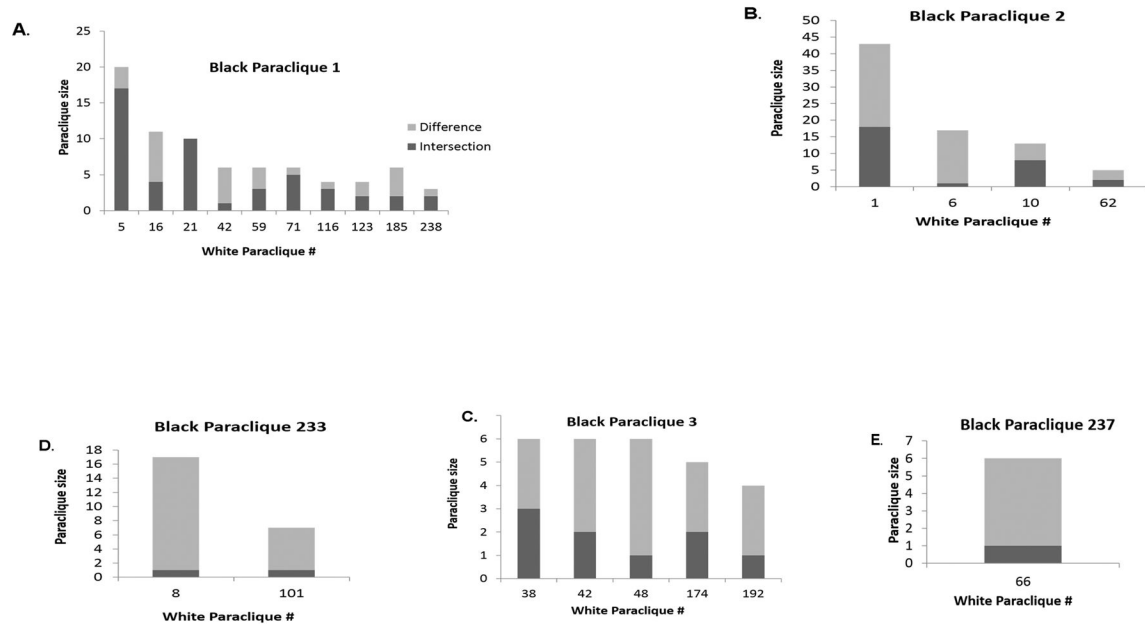


Figure 1. Molecular networks (paracliques) identified in the black population are distributed among several paracliques identified in the white population

Paracliques were computed from transcriptomic data obtained in black and white patients with community-acquired pneumonia at ED presentation. Five paracliques identified in the black population that were comprised of genes involved in apoptosis, cell cycle regulation, lymphocyte development, and metal transport pathways. The distribution of the genes in these paracliques among paracliques identified in the white population was examined. A.) Paraclique #1 had a density of 0.999074, contained 81 genes that were involved in cell cycle and regulation signaling, metallothionein 1E binding. B.) Paraclique #2 had a density of 0.989247, was comprised of 31 genes that were involved in apoptosis, Interleukin -17D, leptin, iodide transporter signaling. C.) Paraclique #3 had a density of 0.991342 and was comprised of 22 genes involved in lymphocyte development, maturation, and adhesion. D.) Paraclique #233 had a density of 1 and was comprised of 3 genes involved in lipid raft signaling and ferritin and metal ion transport. E.) Paraclique 237 identified in the black population is distributed across 1 paraclique identified in the white population.

Table 1

Social and Clinical Demographics of Propensity-Matched Black and White Patients with Community-acquired pneumonia*.

| Clinical Characteristics | Black subjects (N=22) | White subjects (N=22) | P value |
|---|-----------------------|-----------------------|---------|
| Demographics | | | |
| Age, mean (median, SD) | 52.5 (50, 15.8) | 50.1 (47, 16.9) | 0.6276 |
| Sex, female, n (%) | 10 (43%) | 10 (43%) | |
| Illness Severity | | | |
| Charlson comorbidity score, mean | 2.09 | 2.09 | 0.2116 |
| Day 1 APACHE III score ^a , mean (median, SD) | 35 (34,14) | 38 (36, 8) | 0.4279 |
| Need for mechanical ventilation, n (%) | 1 (4%) | 1 (4%) | |
| Required ICU stay, n (%) | 1 (4%) | 1 (4%) | |
| Length of hospital stay, mean (median, SD) | 6.4 (5, 5.3) | 5.8 (5, 4.2) | 0.6192 |

The propensity score was calculated based on age, sex, and pack of cigarettes smoked per day. Black and white patients with community-acquired pneumonia that had the same propensity scores (1:1 match) were paired together.

^aAPACHE III-Acute Physiology and Chronic Health Evaluation II. SE standard error.

Table 2

Population-specific Paraclique: Black Patients Diagnosed with Community-acquired pneumonia

| GENE | FUNCTION |
|--|--|
| C17ORF66; chromosome 17 open reading frame 66 | |
| CENPE; centromere protein E, 312kDa | Plays a key role in the movement of chromosomes toward the metaphase plate during mitosis |
| KIF20A; kinesin family member 20A | Required for chromosome passenger complex (CPC)-mediated cytokinesis. |
| E2F7; E2F transcription factor 7 | Along with E2F8, inhibitor of E2F-dependent transcription that is important for the control of the E2F1-TP53 apoptotic pathway. Directly represses E2F1 transcription. |
| IGLL3; immunoglobulin lambda-like polypeptide 3, pseudogene | |
| PBK; PDZ binding kinase | Phosphorylates MAP kinase p38. Only in mitosis. May play a role in the activation of lymphoid cells. |
| PHGDH; phosphoglycerate dehydrogenase | |
| TMEM155; transmembrane protein 155 | |
| ADAM7; ADAM metallopeptidase domain 7 | May play an important role in sperm maturation and gonadotrope function. A non-catalytic metalloprotease-like protein |
| FANCG; Fanconi anemia, complementation group G | DNA repair protein may operate in a postreplication repair of a cell cycle checkpoint function. May be implicated in interstrand DNA cross-link repair and in the maintenance of normal chromosome stability. Candidate tumor suppressor gene |
| NEK2; NIMA (never in mitosis gene a)-related kinase 2 | Involved in the control of centrosome separation and bipolar spindle formation in mitotic cells and chromatin condensation in meiotic cells |
| SUPT16H; suppressor of Ty 16 homolog (S. cerevisiae) | Component of FACT complex, mRNA elongation, DNA replication and DNA repair. Involved in vitamin D-coupled transcription regulation via its association with the WINAC complex, a chromatin-remodeling complex recruited by vitamin D receptor. |
| GZMA; granzyme A (granzyme 1, cytotoxic T-lymphocyte-associated serine esterase 3) | necessary for target cell lysis in cell-mediated immune responses. Involved in apoptosis |
| GZMK; granzyme K (granzyme 3; tryptase II) | Member of a group of related serine proteases from the cytoplasmic granules of cytotoxic lymphocytes. Cytolytic T lymphocytes (CTL) and natural killer (NK) cells share the remarkable ability to recognize, bind, and lyse specific target cells. |
| GZMH; granzyme H (cathepsin G-like 2, protein h-CCPX) | probably necessary for target cell lysis in cell-mediated immune responses |
| DDIT4; DNA-damage-inducible transcript 4 | Inhibits cell growth by regulating the TOR signaling pathway upstream of the TSC1-TSC2 complex and downstream of AKT1. Promotes neuronal cell death |

Table 3

Population-specific Paraclique: White Patients Diagnosed with Community-acquired pneumonia

| GENE | FUNCTION |
|--|---|
| C15ORF2; chromosome 15 open reading frame 2 | May be involved in spermatogenesis |
| DMWD; dystrophia myotonica, WD repeat containing | May have a regulatory function in meiosis |
| GPR156; G protein-coupled receptor 156 | Orphan receptor |
| MAGEB2; melanoma antigen family B, 2 | Localized in dosage-sensitive sex reversal critical region. Expressed in testis and placenta, and tumors |
| PLN; phospholamban | Postulated to regulate activity of calcium pump of cardiac sarcoplasmic reticulum |
| NUMBL; numb homolog (Drosophila)-like | Negative regulator of NF-kappa-B signaling pathway. The inhibition of NF-kappa-B activation is mediated at least in part, by preventing MAP3K7IP2 to interact with polyubiquitin chains of TRAF6 and RIPK1 and by stimulating the 'Lys-48'-linked polyubiquitination and degradation of TRAF6 in cortical neurons. Role in the process of neurogenesis. |
| TMCO5A; transmembrane and coiled-coil domains 5A | |
| CXCL3; chemokine (C-X-C motif) ligand 3 | Ligand for CXCR2 has chemotactic activity for neutrophils. May play a role in inflammation and exert its effects on endothelial cells in an autocrine fashion. |
| MED31; mediator complex subunit 31 | Coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes. |
| FGF13; fibroblast growth factor 13 | Probably involved in nervous system development and function |

Table 4

Population-specific Paraclique: White Patients Diagnosed with Community-acquired pneumonia

| GENE | FUNCTION |
|---|--|
| PDSS1; Decaprenyl-diphosphate synthase subunit 1 | Supplies decaprenyl diphosphate, precursor for side chain of the isoprenoid quinones ubiquinone-10 |
| EPAS1; endothelial PAS domain protein 1 | Involved in induction of oxygen regulated genes. Binds to core DNA sequence within the hypoxia response element. Regulates VEGF expression and implicated in the development of blood vessels / tubular system of lung. |
| APOO; apolipoprotein O | Promotes cholesterol efflux from macrophage cells. Detected in HDL, LDL and VLDL. May be involved in myocardium-protective mechanisms against lipid accumulation |
| ALCAM; activated leukocyte cell adhesion molecule | Binds to CD6. Involved in neurite extension by neurons. May play a role in the binding of T- and B-cells to activated leukocytes |
| C9ORF30; chromosome 9 open reading frame 30 | |
| DLEC1; deleted in lung and esophageal cancer 1 | May act as a tumor suppressor by inhibiting cell proliferation |
| HBEGF; heparin-binding EGF-like growth factor | Mediates its effects via EGFR, ERBB2 and ERBB4. Required for normal cardiac valve formation and normal heart function. Promotion smooth muscle cell proliferation. May be involved in macrophage-mediated cellular proliferation. It is mitogenic for fibroblasts. |
| OLFML2B; olfactomedin-like 2B | |
| PER2; period homolog 2 (Drosophila) | Negative element in the circadian transcriptional loop. |
| SFMBT1; Scm-like with four mbt domains 1 | |
| THBS1; thrombospondin 1 | Adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix interactions. Binds heparin. Ligand for CD36 mediating antiangiogenic properties |

Table 5

Population-specific Paraclique: White Patients Diagnosed with Community-acquired pneumonia

| GENE | FUNCTION |
|---|--|
| CCRN4L; carbon catabolite repression 4-like | Component of the circadian clock |
| CRISP3; cysteine-rich secretory protein 3 | CRISP-3 in peroxidase-negative granules of neutrophils, in granules of eosinophils, and in exocrine secretions |
| FUT4; fucosyltransferase 4 (alpha (1,3) fucosyltransferase, myeloid-specific) | May catalyze alpha-1,3 glycosidic linkages involved in the expression of Lewis X/SSEA-1 and VIM-2 antigens |
| IGFBP2; insulin-like growth factor binding protein 2 | Inhibits IGF-mediated growth and developmental rates |
| KIAA1024 | |
| KLHDC7B; kelch domain containing 7B | |
| PRG2; proteoglycan 2 | Induces non-cytolytic histamine release from human basophils. Antiparasitic defense and immune hypersensitivity reactions |
| PRSS21; protease, serine, 21 | May regulate proteolytic events associated with testicular germ cell maturation |
| PTP4A3; protein tyrosine phosphatase type IVA, member 3 | Stimulates progression from G1 into S phase. Enhances cell proliferation, cell motility and invasion activity M |
| STOX2; storkhead box 2 | |
| PRSSL1/KLK10; kallikrein-related peptidase 10 | Has a tumor-suppressor role for NES1 in breast and prostate cancer |
| CEACAM8 aka CD67; carcinoembryonic antigen-related cell adhesion molecule 8n | Expressed on granulocytes and leukemic cells |
| LPAR4; lysophosphatidic acid receptor 4 | Transduces a signal by increasing the intracellular calcium ions and by stimulating adenylyl cyclase activity |
| ITPKA; inositol-trisphosphate 3-kinase A | Regulates inositol phosphate metabolism by phosphorylation of second messenger inositol 1,4,5-trisphosphate to Ins(1,3,4,5)P4. |