RNA-seq: primary cells, cell lines and heat stress

Carl J. Schmidt[a], Elizabeth M. Pritchett[a] ,Liang Sun[a], Richard V.N. Davis[a], Allen Hubbard[a], Kalmia E. Kniel[a], Sarah M. Markland[a], Qing Wang[a], Chris Ashwell[b], Michael Persia[c], Max F. Rothschild[d], Susan J. Lamont[d]

[a] Department of Animal and Food Sciences, University of Delaware, Newark DE. 19716
Corresponding author: Carl J Schmidt schmidtc@udel.edu 302-831-1334;Fax 302-831-2822

Elizabeth M. Pritchett[a] 302-831-3102 Fax 302-831-2822
Liang Sun 302-831-3102 Fax 302-831-2822
Richard V.N. Davis 302-831-3102 Fax 302-831-2822
Allen Hubbard 302-831-3102 Fax 302-831-2822
Kalmia E. Kniel 302-831-6513 Fax 302-831-2822
Sarah M. Markland 302-831-6513 Fax 302-831-2822
Qing Wang 302-831-6513 Fax 302-831-2822

[b] Department of Poultry Science, 147 Scott Hall, NCSU, Raleigh, NC 27696
Chris Ashwell 919-627-7337 FAX 919-627-7338

[c] Department of Animal and Poultry Science, 3060 Litton-Reaves Hall, Virginia Tech, Blacksburg, VA. 24061.
Michael Persia 540-231-0735 Fax: 540-231-3010

[d] Department of Animal Science, Iowa State University, 2255 Kildee Hall, Ames IA 50011.
Max F. Rothschild 515-294-6202; Fax: 515-294-2401
Susan J. Lamont 515-294-4100; Fax: 515-294-2401

Transcriptome analysis by RNA-seq has emerged as a high-throughput, cost-effective means to evaluate the expression pattern of genes in organisms. Unlike other methods, such as microarrays or quantitative PCR, RNA-seq is a target free method that permits analysis of essentially any RNA that can be amplified from a cell or tissue. At its most basic, RNA-seq can determine individual gene expression levels by counting the number of times a particular transcript was found in the sequence data. Transcript levels can be compared across multiple samples to identify differentially expressed genes and infer differences in biological states between the samples. We have used this approach to examine gene expression patterns in chicken and human cells, with particular interest in determining response to heat stress.

Three separate *in vitro* cell culture experiments were conducted using the white leghorn chicken hepatocarcinoma LMH cell line, the primary chicken liver (e15) line and the human colorectal cancer epithelial Caco-2 cell line. Cells were freshly plated at subconfluent density and grown for 18 hours at 37°C. Six biological replicates were prepared for the LMH cells while three biological replicates were prepared for the primary chicken hepatocytes and Caco-2 cells. After 18 hours, half of the plates were transferred to 43°C for 2.5 hours after which RNA was isolated, Illumina RNA-seq libraries prepared (Coble et al. 2014; Sun et al. 2015) and sequenced to a depth of 10 million or more reads by the University of Delaware DNA sequencing facility. In addition, we prepared RNA-seq libraries were prepared from six freshly isolated 7-day old chicken livers.

Sequence data were analyzed using the Tuxedo software suite (Gal4 reference build and annotation) and differentially enriched genes identified as indicated below. Signaling and metabolic pathways affected by enriched genes were identified using Gallus Reactome[+] (http://gallus.reactome.org/) and eGIFT (Tudor et al. 2010).

RNA-seq data were combined using chicken:human orthologs to merge the data (Burt et al. 2009) yielding a total of 5333 pairs. Initially, hierarchical clustering was used to compare overall gene expression patterns across the cells and tissues (Fig.1). Although prepared at two different developmental time points (e15 vs. d7), the gene expression patterns of the primary liver culture and freshly prepared liver samples cluster together. The chicken LMH and human Caco-2 cells segregate into a second cluster. Overall, the clustering indicates that the primary cultured liver cells are more similar to fresh liver samples than the established cell lines. The clustering of the cell lines likely reflects the transformed nature and long-term culturing of the LMH and Caco-2 lines. Comparing iTerms (Tudor et al. 2010) for genes enriched in either the cell line samples or the liver samples (Fig.2) indicates enrichment for genes involved in cell proliferation in the cell lines. In contrast, the primary liver and cultured liver samples are enriched for genes involved in functions associated with primary hepatic metabolic processes such as coagulation, cholesterol metabolism, and bile production. The latter observation suggests also that the short-term *in vitro* cultured liver cells may model responses of the intact liver.

Heat Stress

One of the objectives of our work is identifying pathways responsive to heat stress, with a particular interest in identifying evolutionarily conserved responses. To this end, genes were identified that were responsive to heat stress in the cultured liver, LMH and Caco-2 cells. The genes were then mapped to Gallus Reactome[+] (Sun, manuscript in preparation) to identify signaling and metabolic pathways that were modulated (p value <0.05) by heat stress in all three cell types. A total of 280 genes mapped to Gallus Reactome[+] and affected 22 pathways (Table 1). These pathways impact many cellular activities, including the immune system, basic metabolism, response to growth factors and other stimuli, membrane transport of small molecules along with the integrity of the extracellular matrix. Presumably, these represent a core set of pathways that are fundamental to the overall heat stress response.

To identify a conserved set of genes that were responsive to heat stress in all three cultured cell types, pairwise comparisons were conducted of the three experiments to identify genes that were regulated in the same direction (up or down) whose difference was statistically significant (p value < 0.05) (Table 2). The transcription levels of 17 genes were differentially regulated by heat stress across all three cell types; 16 were up-regulated by heat stress while one was down-regulated. Eleven of the up-regulated genes function directly as molecular chaperones or as co-chaperones (HSPB8, SERPINH1, BAG3, HSPH1, HSP90AA1, HSPA5, DNAJA1, CHORDC1, HSPA4, HSPA4L and possibly,

TPPP). Several of the genes products, such as HSP90AA1, HSPA4, and HSPB8 serve as chaperones for multiple client proteins with broad effects on cellular pathways. Some of these gene products function in specific responses. For example HSPA5 functions during endoplasmic reticulum stress, while SERPINH1 affects collagen folding and assembly. CHORDC1 binds the nucleotide-binding domain of HSP90 when the ADP: ATP ratio is high suggesting CHORDC1 may modulate HSP90 activity as a function of energy balance (Gano and Simon 2010).

The hypothesis that TPPP may function as a chaperone or co-chaperone arises from several observations. One observation is from this work showing that TPPP transcription is induced in chicken and human cells by heat stress. Furthermore, the TPPP gene product is known to modulate the tubulin network, promoting tubulin polymerization, microtubule acetylation and bundling, along with alpha-synuclein autophagy (Ejlerskov et al. 2013; Tirian et al. 2003; Tokesi et al. 2010). Taken together this evidence suggests that TPPP may function as a chaperone of proteins that interact with the tubulin matrix. While predominately studied in neuronal cells, TPPP may have similar functions in other cells.

Other products encoded by these heat responsive genes may not function as chaperones. OTUD1 is a deubiquitinase that regulates the level of type 1 interferon in human cells and functions as a suppressor of the innate immune response (Kayagaki et al. 2007). The deubiquitinase activity may allow UTUD1 to serve as a broad regulator of the ubiquitin protein degradation pathway during the unfolded protein response of heat stress. TRA2 is a nuclear protein,

originally identified as controlling sex specific splicing in Drosophila (Amrein et al. 1990; Baker et al. 1989; Goralski et al. 1989) and controls alternative splicing in vertebrates (Zhang et al. 2007). Possibly, heat stress induction of TRA2 leads to an evolutionarily conserved pattern of alternative splicing (Strasburg and Chiang 2009).

Three of the conserved heat responsive genes function either extracellularly or by interaction with the extracellular matrix. OTOA was originally identified as a protein that links the apical surface of inner ear epithelial cells to the extracellular gel (Zwaenepoel et al. 2002). IMPG2 encodes a hyaluronic binding proteoglycan (Acharya et al. 2000). The final heat responsive up-regulated gene, SNTB2 encodes a member of the syntrophin family that links the cytoskeleton with the extracellular matrix (Ervasti and Campbell 1993). The sole gene down regulated by heat stress across all three cell types was VRK1, Vaccinia related kinase 1. This kinase functions during the cell cycle, and is required for transiting the G1 phase (Valbuena et al. 2008). Down-regulation by heat would likely reduce the rate of cell proliferation, possibly permitting more time to repair stress associated cell damage.

The great depth of reads provided by RNA-seq can challenge assertions that a gene is only expressed in a specific cell or tissue. For example the OTOA gene was originally identified as expressed only in inner ear, while the IMPG2 gene has been identified as specific to retinal and pineal tissue (Acharya et al. 2000). However, we detect low levels of both OTOA and IMPG2 transcripts in all three cell types. Inspection of the RNA-seq coverage for both genes revealed multiple

reads mapping to the corresponding transcripts with the preponderance coming from exons (Fig. 3). Also, some of the reads span an intron. Taken together these observations suggest that the detected reads arise from mature transcripts rather than precursor RNAs or genomic contamination. The conclusion is that the reads from these genes in our experiments arose from true OTOA or IMPG2 transcripts. However, there is no current evidence that the OTOA or IMPG2 transcripts give rise to protein. This demonstrates an important current challenge: linking data from high throughput sequencing and proteomics analysis to verify translation of such mRNAs.

Future Directions:

A major application of RNA-seq transcriptome is determining gene expression profiles. Currently, many such studies focus on identifying genes based upon the annotated genomic sequence for the target organism. However, RNA-seq data can provide a rich source of data for transcripts that are currently not recognized in a genome's annotation file (Smith et al. current issue). Combining data from short-read (such as data generated by Illumina) and long-read (PacBio) RNA-seq technologies can provide an even better understanding of transcript structure, significantly improving genome annotation. Alternative splicing, allele specific expression, microRNAs and lncRNAs can all be identified with RNA-seq data, providing a rich catalog of the diversity of transcripts found in samples. Given the (relative) ease with which this data can be collected, an important current need is improving and extending bioinformatics tools. While many such tools exist, the community needs to work together to develop intuitive web-based bioinformatics

pipelines and platforms that are appropriate to users with varying levels of computer sophistication. Ultimately, combining an incredibly rich set of sequence data with user-friendly bioinformatics tools will provide complete gene expression profiles of organisms in many different biological states.

References:

Acharya S, Foletta VC, Lee JW, Rayborn ME, Rodriguez IR, Young WS, 3rd, Hollyfield JG: Spacrcan, a novel human interphotoreceptor matrix hyaluronan-binding proteoglycan synthesized by photoreceptors and pinealocytes. The Journal of Biological Chemistry 275:6945-6955 (2000).

Amrein H, Maniatis T, Nothiger R: Alternatively spliced transcripts of the sex-determining gene TRA2 of Drosophila encode functional proteins of different size. The EMBO Journal 9:3619-3629 (1990).

Baker BS, Burtis K, Goralski T, Mattox W, Nagoshi R: Molecular genetic aspects of sex determination in drosophila melanogaster. Genome / National Research Council Canada = Genome / Conseil national de recherches Canada 31:638-645 (1989).

Burt DW, Carre W, Fell M, Law AS, Antin PB, Maglott DR, Weber JA, Schmidt CJ, Burgess SC, McCarthy FM: The chicken gene nomenclature committee report. BMC genomics 10 Suppl 2:S5 (2009).

Coble DJ, Fleming D, Persia ME, Ashwell CM, Rothschild MF, Schmidt CJ, Lamont SJ: RNA-seq analysis of broiler liver transcriptome reveals novel responses to high ambient temperature. BMC Genomics 15:1084 (2014).

Ejlerskov P, Rasmussen I, Nielsen TT, Bergstrom AL, Tohyama Y, Jensen PH, Vilhardt F: Tubulin polymerization-promoting protein (TPPP/p25alpha) promotes unconventional secretion of alpha-synuclein through exophagy by impairing autophagosome-lysosome fusion. The Journal of Biological Chemistry 288:17313-17335 (2013).

Ervasti JM, Campbell KP: A role for the dystrophin-glycoprotein complex as a transmembrane linker between laminin and actin. The Journal of Cell Biology 122:809-823 (1993).

Gano JJ, Simon JA: A proteomic investigation of ligand-dependent HSP90 complexes reveals chordc1 as a novel adp-dependent hsp90-interacting protein. Molecular & Cellular proteomics : MCP 9:255-270 (2010).

Goralski TJ, Edstrom JE, Baker BS: The sex determination locus Transformer-2 of Drosophila encodes a polypeptide with similarity to RNA binding proteins. Cell 56:1011-1018 (1989).

Kayagaki N, Phung Q, Chan S, Chaudhari R, Quan C, O'Rourke KM, Eby M, Pietras E, Cheng G, Bazan JF, Zhang Z, Arnott D, Dixit VM: Duba: A deubiquitinase that regulates type i interferon production. Science 318:1628-1632 (2007).

Strasburg GM, Chiang W: Pale, soft, exudative turkey--the role of ryanodine receptor variation in meat quality. Poultry Science 88:1497-1505 (2009).

Sun L, Lamont SJ, Cooksey AM, McCarthy F, Tudor CO, Vijay-Shanker K, DeRita RM, Rothschild MF, Ashwell C, Persia ME, Schmidt CJ: Transcrptome response to heat stress in a chicken hepatocellular carcinoma cell line. Poultry Science In press (2015).

Tirian L, Hlavanda E, Olah J, Horvath I, Orosz F, Szabo B, Kovacs J, Szabad J, Ovadi J: TPPP/p25 promotes tubulin assemblies and blocks mitotic spindle formation. Proceedings of the National Academy of Sciences of the United States of America 100:13976-13981 (2003).

Tokesi N, Lehotzky A, Horvath I, Szabo B, Olah J, Lau P, Ovadi J: TPPP/p25 promotes tubulin acetylation by inhibiting histone deacetylase 6. The Journal of Biological Chemistry 285:17896-17906 (2010).

Tudor CO, Schmidt CJ, Vijay-Shanker K: eGIFT: Mining gene information from the literature. BMC Bioinformatics 11:418 (2010).

Valbuena A, Lopez-Sanchez I, Lazo PA: Human VRK1 is an early response gene and its loss causes a block in cell cycle progression. PloS One 3:e1642 (2008).

Zhang X, Moor AN, Merkler KA, Liu Q, McLean MP: Regulation of alternative splicing of liver scavenger receptor class B gene by estrogen and the involved regulatory splicing factors. Endocrinology 148:5295-5304 (2007).

Zwaenepoel I, Mustapha M, Leibovici M, Verpy E, Goodyear R, Liu XZ, Nouaille S, Nance WE, Kanaan M, Avraham KB, Tekaia F, Loiselet J, Lathrop M, Richardson G, Petit C: Otoancorin, an inner ear protein restricted to the interface between the apical surface of sensory epithelia and their overlying acellular gels, is defective in autosomal recessive deafness DFNB2. Proceedings of the National Academy of Sciences of the United States of America 99:6240-6245 (2002).

Tables

Gallus Reactome plus Pathways
Adaptive Immune System
Biological oxidations
Collagen formation
Degradation of the extracellular matrix
Generic Transcription Pathway
Innate Immune System
Ion channel transport
Metabolism of amino acids and derivatives
Metabolism of carbohydrates
Metabolism of lipids and lipoproteins
Peptide hormone metabolism
Platelet activation, signaling and aggregation
Post-translational protein modification
Regulation of Insulin-like Growth Factor (IGF) Transport and Uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)
Regulation of mRNA Stability by Proteins that Bind AU-rich Elements
Signaling by ERBB2
Signaling by ERBB4
Signaling by GPCR
Signaling by NOTCH
Signaling by NGF
SLC-mediated transmembrane transport
Transmission across Chemical Synapses

Table 1. Pathways affected by heat stress in *in vitro* cultured primary cultured liver, LMH and Caco-2 cells.

RNA-seq and Stress

Symbol	Description	TC_Liver P-value	Caco-2 P-value	LMH P-value	Mean Control TC_Liver	Mean Heat TC_Liver	Mean Control Caco-2	Mean Heat Caco-2	Mean Control LMH	Mean Heat LMH
HSPA5	Heat Shock 70kDa Protein 5	0.0002	0.0088	<.0001	418	2411	265	562	316	1710
HSPB8	Heat Shock 22kDa Protein 8	0.0004	0.0145	0.0001	1.17	12.4	19.1	52.7	41.2	1484
HSP90AA1	Heat Shock Protein 90kDa alpha 1	<.0001	0.0423	<.0001	52.4	549	239	943	205	1369
BAG3	BCL2-associated athanogene 3	0.0002	0.0307	<.0001	47.6	665	14.8	198	33.8	806
HSPH1	Heat Shock 105kDa/110kDa Protein 1	<.0001	0.0180	<.0001	12.6	183	43	468	47.9	373
TRA2A	Transformer 2 Alpha	0.0002	0.0135	<.0001	36.5	68.9	33	63.8	79.9	194
HSPA4L	Heat Shock 70kDa Protein 4	<.0001	0.0417	0.0003	30.5	91.3	2.84	33.9	78.7	175
DNAJA1	DnaJ (Hsp40)	<.0001	0.0099	<.0001	28.3	114	161	509	54.1	171
CHORDC1	Cysteine and Histidine-rich domain containing 1	0.0009	0.0366	<.0001	7.76	29.6	62	115	39.4	124
OTUD1	OTU Domain Containing 1	0.0055	0.0146	0.0021	15.1	164	8.03	21.3	14	104
HSPA4	Heat Shock 70kDa Protein 4	<.0001	0.0342	<.0001	18.5	66.7	65.4	117	29.1	88.8
TPPP	Tubulin Polymerization Promoting Protein	0.0064	0.0451	0.0076	0.42	0.89	0.09	0.43	11.6	26.6
SNTB2	Syntrophin, beta 2	0.0018	0.0072	0.0014	2.29	4.93	4.98	9.81	6.97	15.4
VRK1	Vaccinia Related Kinase 1	<.0001	0.0106	0.0010	5.54	2.03	22.2	6.79	24.8	7.23
SERPINH1	Serpin Peptidase Inhibitor, clade H -1	0.0005	0.0346	<.0001	31.1	87.8	231	799	0.29	2.41
OTOA	Otoancorin	0.0018	0.0289	<.0001	0.14	0.96	0.02	0.12	0.53	2.19
IMPG2	Interphotoreceptor Matrix Proteoglycan 2	0.0002	0.0315	0.0024	0.11	1.09	0.06	0.12	0.39	1.14

Table 2. Differentially Expressed Genes responsive to heat stress. P-value for difference between heat-stressed and control samples derived from primary cultured liver (TC_Liver), LMH cells (LMH) and Caco-2 cells followed by the mean values for the control and heat stress RPMK values.

Figure Legends:

Figure 1. Hierarchical clustering of RNA-seq data from primary cultured liver cells (TCLiver), fresh Liver, Chicken white-leghorn hepatocarcioma cell line (LMH) and human colorectal cancer cell line (Caco-2).

Figure 2. Stacked Histogram of iTerms describing biological processes affected by genes enriched in the primary cultured liver samples plus fresh liver (red) compared with genes enriched in the LMH plus Caco-2 cell lines (blue).

Figure 3. Track diagrams from UCSC browser plotting chicken LMH RNA-seq reads from the OTOA (A & B) or IMPG2 (C & D). A and C correspond to the entire gene region while B and D have been zoomed into show coverage of individual exons along with reads that span the intron indicated splicing.
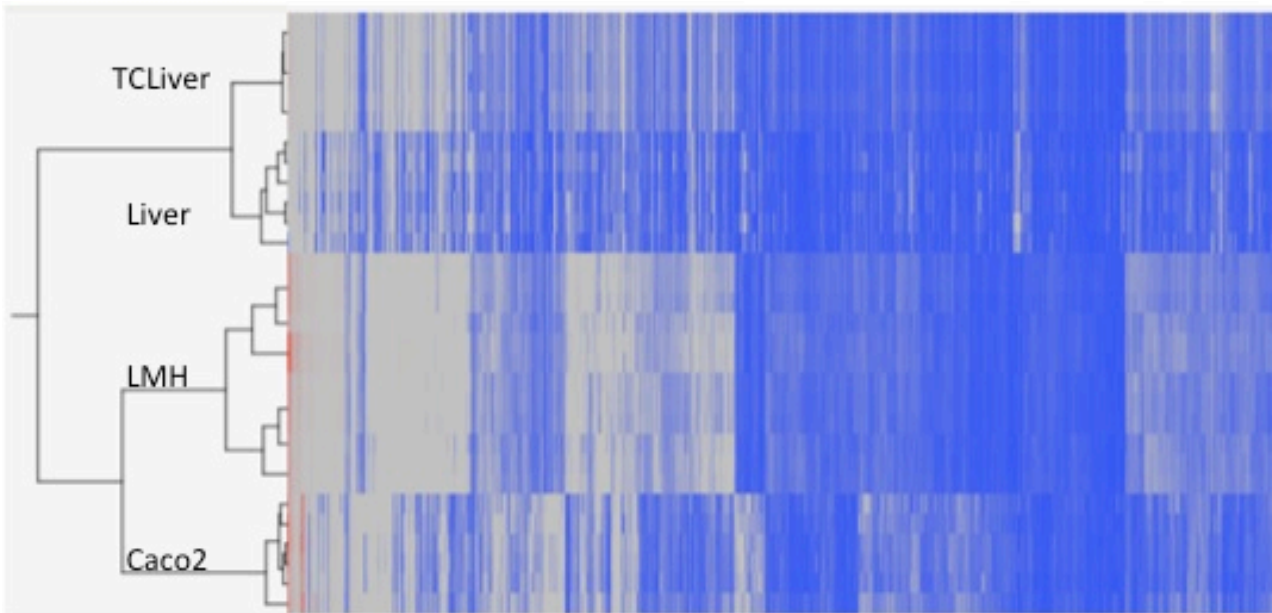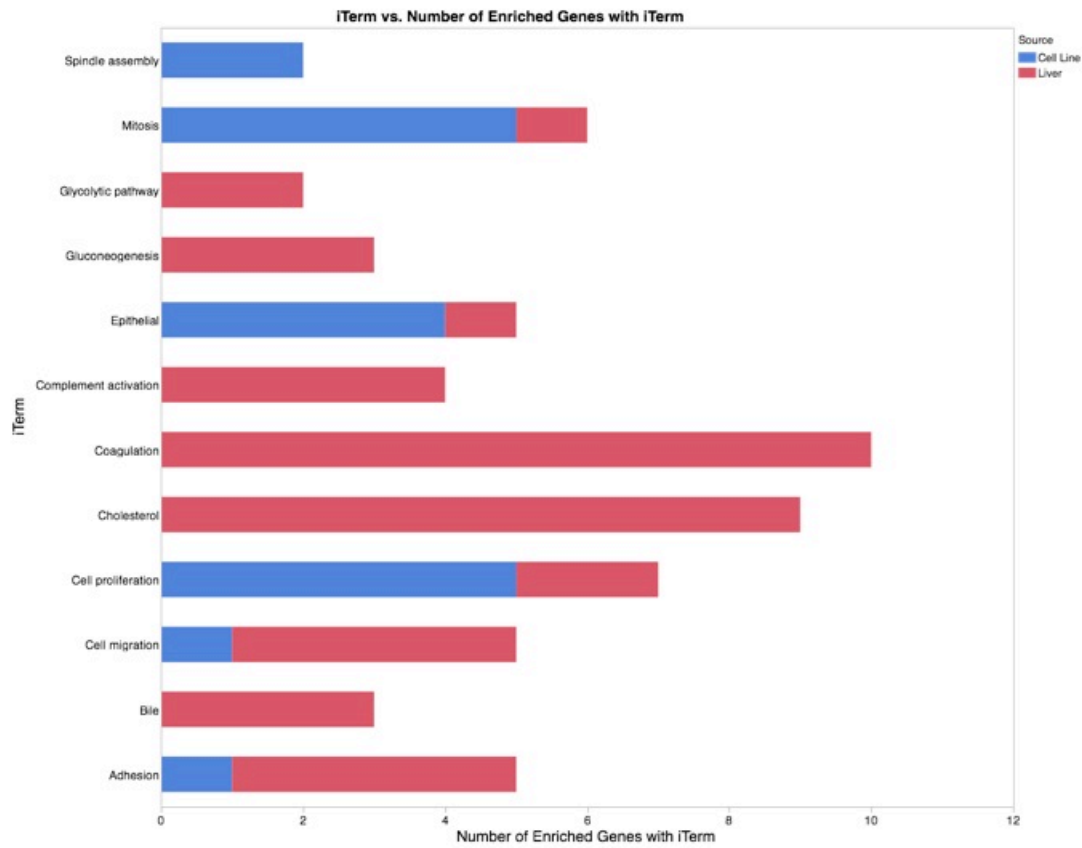
Figure 1. Hierarchical clustering

Figure 2. Stacked Histogram of iTerms

Fig. 3A

Fig. 3 B

Fig. 3C

Fig. 3D