CREST

RNAnet

Technical Highlights

W. B. Langdon

King's College, London

Introduction

- RNAnet is a map of human gene expression
- Derived from 40000 Affymetrix GeneChips (da Silva Camargo) and 30000 human genes defined by Ensembl (Sanchez-Graillet, Rowsell)
- RNAnet contains all 290 million pair-wise correlations.
- Offline analysis based on R and unix scripts
- Concentrate upon PHP, JavaScript, to drive firefox mycoplasma contamination example.

Mycoplasma Contamination

- Affymetrix probeset 1570561_at turns out to measure bacterial contamination not human
  - "Unexpected presence of mycoplasma probes on human microarrays", Biotechniques.
- Use RNAnet to display normalised HG-U133 data from 2700 samples in a few seconds. (Typically experiments use a few microarrays.)

Correlation Heat Map - SeaMonkey

File Edit View Go Bookmarks Tools Window Help

http://bioinformatics.essex.ac.uk/users/wlangdon/rnanet/correlation.html

Search

Enter search term, keyword, or web address

Home Bookmarks KCL Webmail Release Notes Plug-ins Extensions Where?

Correlation of Expression of Human Genes in GEO

ENSE lookup

Examples: ENSE00001045180

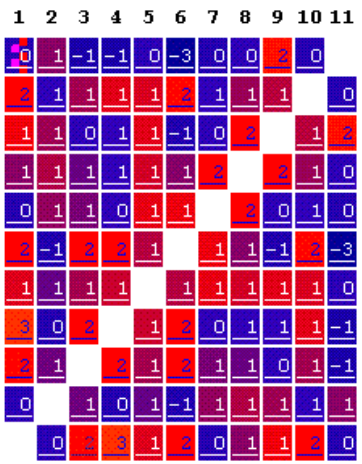
1570561_at

214067_at-497-757:84-60 214067_at-651-391:52-28
214067_at-745-505:37-13
1556291_at
211462_s_at.pm9-11

PM only Heatmap text done Fri, 25 Sep 2009 08:27:51 GMT

Example output

2757 HG_U133_Plus_2 WBL 04 August 2008

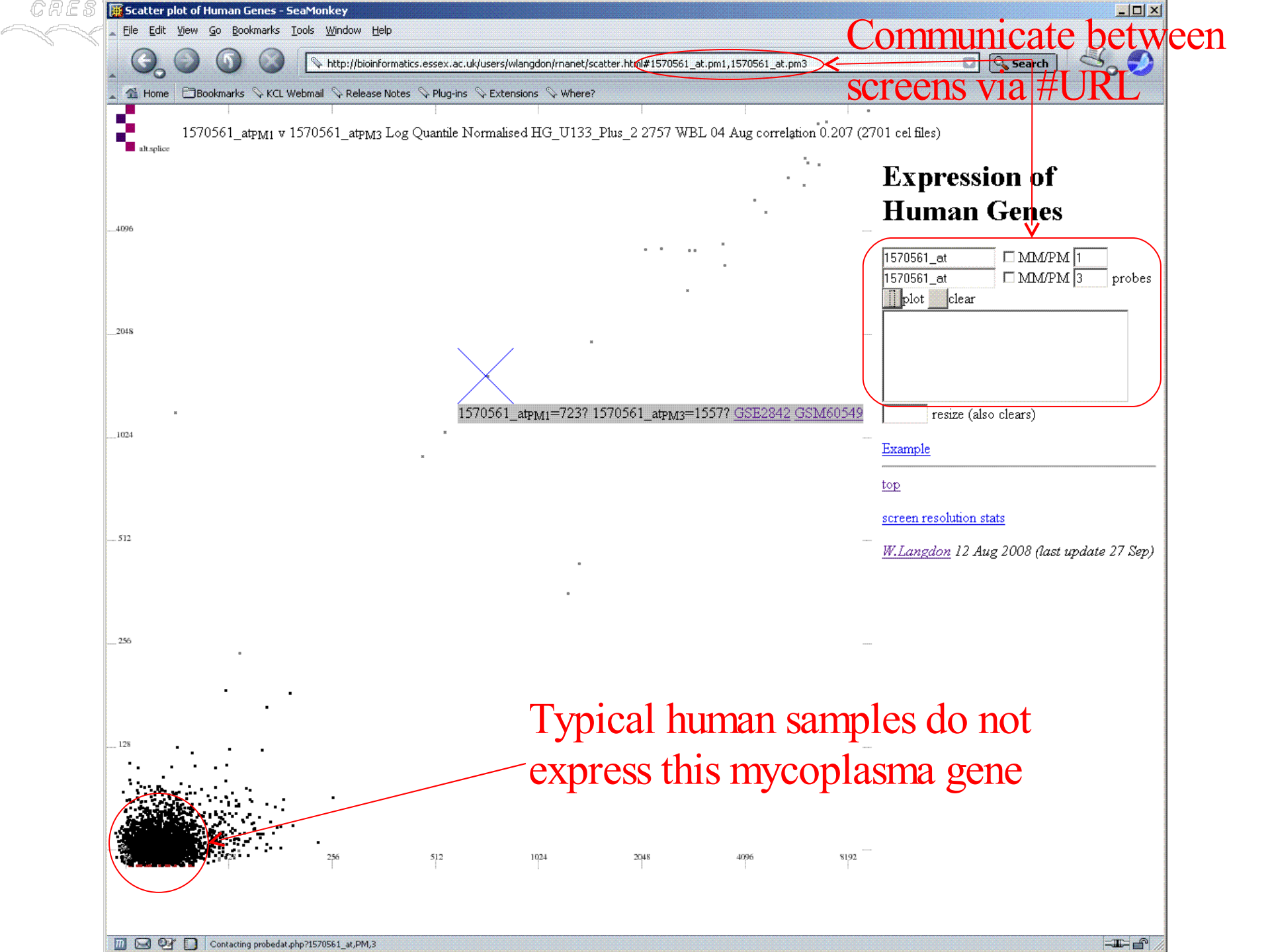


11	1570561_at.pm11	1024,291	207	AATTGAAACATCTCAGTAGCAGCAG	74	1.16	0	1	-1	-1	0	-3	0	0	2	0						
10	1570561_at.pm10	1027,601	196	GACACGCTGTGAATTGAAACATCTC	73	1.14	2	1	1	1	1	2	1	1	1		0					
9	1570561_at.pm9	896,483	184	GCCGAATGATGAGACACGCTGTGAA	87	1.18	1	1	0	1	1	-1	0	2		1	2					
8	1570561_at.pm8	948,409	171	CGGTGAATCCATAGCCGAATGATGA	73	1.13	1	1	1	1	1	1	2		2	1	0					
7	1570561_at.pm7	552,923	162	TGTCATAATCGGTGAATCCATAGCC	68	1.11	0	1	1	0	1	1		2	0	1	0					
6	1570561_at.pm6	431,807	148	GGGTAATGCCTAATTGTCATAATCG	122	1.31	2	-1	2	2	1		1	1	-1	2	-3					
5	1570561_at.pm5	282,275	104	AATGAGCATTTAGACGGAGATTTCC	70	1.12	1	1	1	1		1	1	1	1	1	0					
4	1570561_at.pm4	321,517	88	GCTTCGTTTAGCTGGAAATGAGCAT	72	1.15	3	0	2		1	2	0	1	1	1	-1					
3	1570561_at.pm3	302,1121	75	TTACCTGCGATAAGCTTCGTTTAGC	68	1.12	2	1		2	1	2	1	1	0	1	-1					
2	1570561_at.pm2	672,855	67	GGACGTGATTACCTGCGATAAGCTT	77	1.15	0		1	0	1	-1	1	1	1	1	1					
1	1570561_at.pm1	726,261	56	AAGTCGATGAAGGACGTGATTACCT	82	1.19		0	1	3	1	2	0	1	1	2	0					

top

StatCounter.com

Contacting corrdat.php?217712,217713,217714,217715,217716,217717,217718,217719,217720,217721,217722


Scatter plot of Human Genes - SeaMonkey
File Edit View Go Bookmarks Tools Window Help
http://bioinformatics.essex.ac.uk/users/wlangdon/rnanet/scatter.html#1570561_at.pm1,1570561_at.pm3
Search
Home Bookmarks KCL Webmail Release Notes Plug-ins Extensions Where?
Communicate between screens via #URL
1570561_atPM1 v 1570561_atPM3 Log Quantile Normalised HG_U133_Plus_2 2757 WBL 04 Aug correlation 0.207 (2701 cel files)
alt.splice
4096
2048
1024
512
256
128
1570561_atPM1=723? 1570561_atPM3=1557? GSE2842 GSM60549
Expression of Human Genes
1570561_at MM/PM 1
1570561_at MM/PM 3 probes
plot clear
resize (also clears)
Example
top
screen resolution stats
W.Langdon 12 Aug 2008 (last update 27 Sep)
Typical human samples do not express this mycoplasma gene
Contacting probedat.php?1570561_at,PM,3

scatter.html

Communicate between screens via #URL

<body initialise_scatter();
if (event.persisted) initialise_scatter(); >

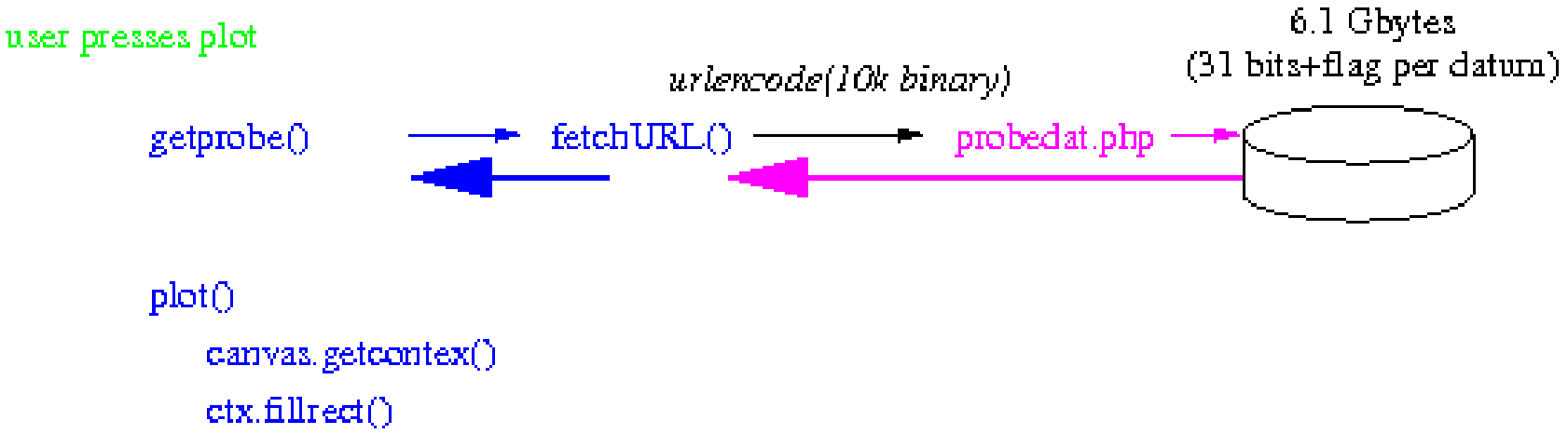
<canvas

</canvas>

```
<form name="probes">
<input type=text name="X_name" size="16" autocomplete="OFF" ONCHANGE="saveprobes()">
<input type=checkbox name="X_MM" style="background-color:transparent;" ONCHANGE="saveprobes()">MM/PM
<input type=text name="X_num" size="2" autocomplete="OFF" ONCHANGE="saveprobes()">
<br>
<input type=text name="Y_name" size="16" autocomplete="OFF" ONCHANGE="saveprobes()">
<input type=checkbox name="Y_MM" style="background-color:transparent;" ONCHANGE="saveprobes()">MM/PM
<input type=text name="Y_num" size="2" autocomplete="OFF" ONCHANGE="saveprobes()">
probes
<br>
<input type=button onclick="scatter()" >plot
<input type=button onclick="clear_scatter()" >clear
<TEXTAREA name="gses"
COLS="30" ROWS="5"
WRAP="SOFT"
autocomplete="OFF"
ONCHANGE="savegses(this)"
>
</TEXTAREA>
<input type=text name="size" size="4" autocomplete="OFF" ONCHANGE="resize(this)">
resize (also clears)
</form>
```

JavaScript to get data and plot it

scatter.js ↔ server (php)



getprobe (JavaScript)

All data from server is cached

Tell user request is active

```
function getprobe(probe,button) {
  if(!(probe in cache)) {
    button.style.background = 'red';
    var url = "probedat.php?"+probe;
    //alert("fetchURL("+url+")");
    var s = fetchURL(url);
    if(s.length<80) {
      alert("Failed to get probe "+probe+"\n"+s.trim()); return;}
    var s = unescape(s);
    var head = s.substr(0,80).split(" ");
    if(head[0]!=s.length-80) {
      alert("fetchURL("+url+") failed 2 length="+s.length+" v "+head[0]+" 80=`"+s.substr(0,80)+"' end=`"+s.substr(2757)+"'"); return;}
    else if(head[0]!=2*ncels) {
      alert(url+" not compatible "+head); return;}
    else if(head[1].toLowerCase()!=probe.toLowerCase()) {
      alert("fetchURL("+url+") returned wrong data "+head[1]+" v "+probe); return;}
    cache[probe]= new Object();
    cache[probe].head=s.substr(0,80);
    cache[probe].dat=new Array((s.length-80)/2);
    cache[probe].ok=new Array((s.length-80)/2);
    for(var i=80;i<s.length;i+=2) {
      var IV0 = s.charCodeAt(i);
      var IV1 = s.charCodeAt(i+1);
      var Int = btoint(IV0,IV1);
      var I = (i-80)/2;
      if(Int<32768) {
        cache[probe].dat[I] = i2toreal(Int);
        cache[probe].ok[I]  = true;
      } else {
        cache[probe].dat[I] = i2toreal(Int-32768);
        cache[probe].ok[I]  = false;
      }
    }
  }
  return cache[probe];
}
```

Check reply matches request

Extract two bytes at a time,
convert to ok flag and real

net.js fetchURL (synchronous)

```
function fetchURL(url) {
//IE or Moz
  defaultStatus = "Requesting XMLHttpRequest";
  var   xmlhttp = new XMLHttpRequest();
   defaultStatus = "Contacting "+url;
   xmlhttp.open("GET", url, false, "", "");
   xmlhttp.send(null);
   var res = xmlhttp.responseText;
   return res;
```

Try and tell user what is happening

Talk to server and *wait* for all of its reply

```
function fetchURL(url) {
//netscape
  defaultStatus = "Contacting "+url;
  var dest = new java.net.URL(url);
  var dis = new java.io.DataInputStream(dest.openStream());
  var res = "";
  while ((line = dis.readLine()) != null) {
    res += line;
    res += java.lang.System.getProperty("line.separator");
  }
  dis.close();
  defaultStatus += "done.";
  return res;
```

probedat.php

```
<?php
//WBL  4 Aug 2008 based on probeset.php r1.10 $Revision: 1.10 $

//WBL  3 Sep 2008 use probelib

require_once('log.php');
require_once('grammar.php');
require_once('probelib.php');

$parts = checkargsandheader(3);
$probeset=strtolower($parts[0]); //"1552573_s_at";
$pm      =(strtoupper($parts[1])=="PM");
$probenum=intval($parts[2]);

$parts = findrec($probeset);
$probeset = $parts[0]; //use affy Upper/lower case
$rec      = $parts[1];
$nprobes  = $parts[2]/2;

if($probenum<1 || $probenum>$nprobes) done("Only $nprobes in $probeset. Probe $probenum missing",41);
$rec = $rec + ($probenum-1) + (($pm)? 0 : $nprobes);

$row = readprobefiles($rec,1);
$ncels = ncelfiles();
$recordsize=strlen($row);
#use dummy for $file_number

echo rawurlencode(firstline($recordsize." ".
                            $probeset.",".
                            (($pm)? "PM":"MM").",".
                            $probenum." ".$nprobes.
                            " Log Quantile Normalised HG_U133_Plus_2 $ncels WBL ".
                            probeversion()).
                  $row);
logx($status,$recordsize,$ncels,$rec,"");
?>
```

probedat.php?1570561_at,PM,1

Look up probeset (binary chop)

Read data (fseek, fread)

Send data back (as text)

Header identifies data and size

Down side - PHP

- PHP `file` access simple to use but limited to files of a few MB
- PHP random access (`fseek`) limited to 2Gbytes (we used 4 files).
- Overhead of generating text from numbers is high (15sec/MB):
  - Also high overhead on user's browser interpreting/displaying big html
  - Have to save binary file and pass its URL to user, but then must manage temporary files/URLs.

Down side - Desktop

- Synchronous fetchURL ok but server requests must be fast.
  - Colouring buttons browser may delay showing change
  - Status line does not work as advertised.
  - Ok for user feedback
- Screens made by JavaScript, forward/backward not as user likes
  - #URL provides way of communicating between screens.
  - Avoid user re-entering data. Keep browser responsive (no heavy CPU).
  - Encourage users to open new tabs instead.
- Cache allows browser to burn memory not network latency
- Use plain text where feasible
- Not multi browser (too labour intensive). Firefox <canvas>

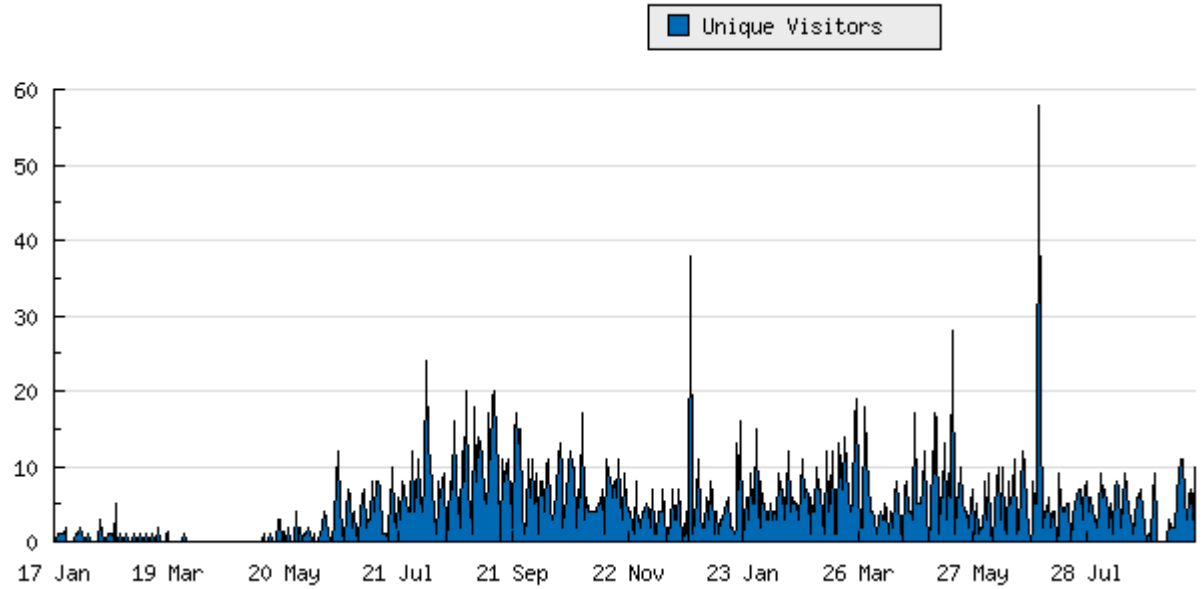
- **Better ways?**

Conclusion

- Web browser offers fast route to user's desktop
  - No installation overhead
- Data are published Affy GeneChips. They cover virtually all medically interesting human tissues.
- RNAnet http://bioinformatics.essex.ac.uk/users/wlangdon/ allows specialist to interactively explore millions of RNA expression data and their correlations.
  - Presented at UKAffy and EMBO 2008. Essex technical report CES-486.

END

Usage



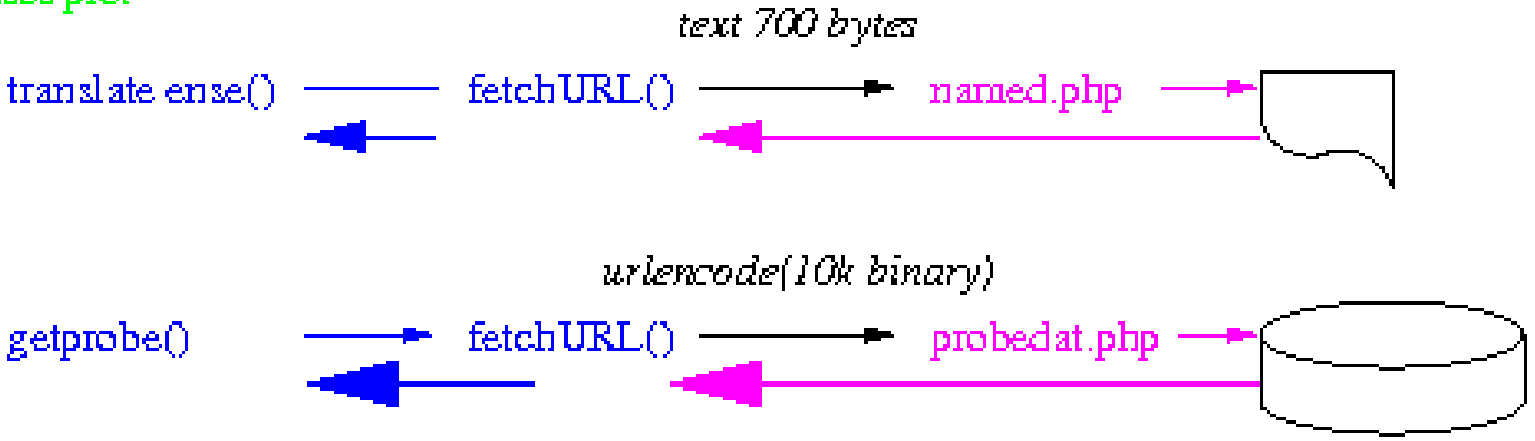
- RNAnet use 362156. More than 96% of use by web spiders.
- Netstat estimates real usage (excluding Essex and King's) 2939

Human Gene Co-Expression

- Correlations between human exons are highly non-random. The network formed by pairs of strongly correlated exons is a "small world":
  - No exon is strongly correlated with all others
  - 2% exons are strongly correlated with >1000 others
  - Most exons are strongly correlated with <48 others
  - By using multiple steps most exons can be reached
  - 12% of exons not strongly correlated with any others
  - Some "small world" power laws found.
- RNAnet is an interactive tool to explore RNA expression (protein gene and non-protein coding) of thousands of published 3' GeneChips covering virtually all medically interesting human tissues.

Converting Ensembl exon id to Affy probeset

user presses plot



plot()

canvas.getcontex()

ctx.fillrect()

namedat.php

```php
<?php
//WBL 24 Aug 2008 based on probedat.php r1.5 $Revision: 1.7 $

//WBL  3 Sep 2008 use probelib

require_once('log.php');
require_once('grammar.php');
require_once('probelib.php');

$parts = checkargsandheader(1);
$probeset=$parts[0];
if(substr(strtoupper($parts[0]),0,4)=="ENSE") {
  $probeset = sprintf("ENSE%011d",substr($parts[0],4));
  $search   = "^$probeset ";
  $rec      = 0;
  $nprobes  = 0;
  $grepfile = "/global1/users/wlangdon/HG-U133_Plus_2_ense.txt";
} else {
if(strlen($probeset)<6) done("probeset name `$probeset' too short",50);

$parts = findrec($probeset);
$probeset = $parts[0]; //use affy Upper/lower case
$search   = "^$probeset ";//tab!
$rec      = $parts[1];
$nprobes  = $parts[2]/2;
$grepfile = "HG-U133-PLUS_probe_tab";
}
#echo "Content-Type: text/plain\n\n";
header("Content-type: text/plain");
echo firstline("-1 ".$probeset." ".$rec." ".$nprobes." HG_U133_Plus_2 WBL ".
               date("d F Y",filemtime($grepfile))."\n";
$command  = "egrep $search $grepfile";
system("ulimit -t 10\n".$command,$status);

logx($status,"","",$rec,$command);
?>
```