Folate system correlations in DNA microarray data

Tomas Radivoyevitch

Department of Epidemiology and Biostatistics

Case Western Reserve University, Cleveland, Ohio 44106 USA

Tel: 216-368-1965; Email: radivot@hal.cwru.edu

Abstract

Background

Gene expression data is abundantly available from the Gene Expression Omnibus (GEO) and various websites. Pathway specific analyses of gene-gene correlations across these datasets remain relatively unexplored, though they could be informative.

Methods

Assuming direct proportionality between enzyme mRNA and protein levels, a mathematical model of folate metabolism is used to reduce the dimensionality of folate gene expression measurements to two dimensions, the predicted DNPS and DNTS metabolic fluxes.

Results

Positive correlations within and between the folate pathways of de novo purine synthesis (DNPS) and de novo thymidylate synthesis (DNTS) are observed in gene expression data. For steady state measurements across childhood leukemia patients, positive correlations between DNPS and DNTS are consistent with higher proliferative fractions requiring higher levels of both fluxes. For cells exposed to ionizing radiation, transient increases in both pathways are consistent with DNA damage driven dNTP demand, and a steadily decreasing backdrop of these fluxes is consistent with radiation induced cell cycle arrest.

Conclusions

Correlations exist within and between the DNPS and DNTS pathways of the folate system.

Background

The folate system (Figure 1A) is central to *de novo* purine synthesis (DNPS) and *de novo* thymidylate synthesis (DNTS) and is a key target of several anti-cancer agents. For example, methotrexate (MTX), in its polyglutamated forms, inhibits dihydrofolate reductase (DHFR), thymidylate synthetase (TS), glycinamide ribonucleotide formyltransferase (GART), and other folate system enzymes (see the MTX containing reaction equations [1] in Methods); the novel multi-targeted anti-folate ALIMTA has similar targets [2], though with a very different spectrum of inhibition constants K_i such that GART inhibition is dominant [3]; the anti-folate raltitrexed (Tomudex) mostly inhibits TS [4]; and 5-fluorouracil also inhibits TS (as FdUMP), though it also kills cells via incorporation into DNA and RNA [5].

Folate has two key double bonds (Figure 1B, arrows). Saturating the upper double bond yields dihydrofolate (DHF) and adding an additional molecule of hydrogen across the lower double bond yields tetrahydrofolate (THF). Folates serve as donors of single carbons in any one of three oxidation states: 5-methyl-THF (CH_3THF; reduced), 5,10 methylene-THF (CH_2THF; intermediate) and 10-formyl-THF (CHOTHF; oxidized). The single carbon donor CH_3THF is used to convert homocysteine into methionine, the donor CH_2THF is used (along with the molecule of hydrogen across the lower double bond) to convert dUMP (deoxyuridylate) into dTMP (thymidylate) and the donor CHOTHF is used to set up ring closure reactions in *de novo* purine synthesis.

The folate model of Morrison and Allegra [1] was developed to simulate MTX responses (Figure 2; Methods). It will be used here to convert multi-dimensional folate gene expression data into DNPS and DNTS metabolite flux predictions. In this conversion direct proportionality is assumed between enzyme mRNA levels and enzyme protein levels [6].

Methods

Mathematical Model of Morrison and Allegra

The folate cycle model of Morrison and Allegra [1] has the following mathematical form:

$$\frac{dDHF^T}{dt} = \sum_{9,12} r_i - \sum_{10,11} r_i \qquad \frac{dTHF}{dt} = \sum_{1R,2R,4,6,7,10} r_i - \sum_{1,2,5} r_i \qquad \frac{d(CH_2THF)}{dt} = \sum_{1,2} r_i - \sum_{1R,2R,3,8,9} r_i$$

$$\frac{d(CH_3THF)}{dt} = r_3 - r_4 \qquad \frac{dCHOTHF}{dt} = \sum_{5,8} r_i - \sum_{6,7} r_i \qquad \frac{dCHODHF}{dt} = r_{11} - r_{12}$$

$$\frac{dHCHO}{dt} = r_{2R} - r_2 \qquad \frac{d(dUMP)}{dt} = U_0 - r_9 \qquad \frac{dGAR}{dt} = G_0 - r_6$$

$$\frac{dFGAR}{dt} = r_6 - r_{13} \qquad \frac{dAICAR}{dt} = r_{13} - r_7 - r_{12}$$

These equations restate the system configuration information of Figure 1A, i.e. they state that the rate at which a metabolite concentration increases equals the sum of the synthesis reaction fluxes (arrows into a node) minus the sum of the degradation reaction fluxes (arrows leaving a node). The r_i in these equations are:

$$r_1 = \frac{V_{SHMT}}{(1+K_{THF}/THF)(1+K_{SER}/Serine)} \qquad r_{1R} = \frac{V_{SHMTR}}{(1+K_{CH_2THF}/CH_2THF)(1+K_{GLY}/Glycine)}$$

$$r_2 = h_p(THF)(HCHO) \qquad r_{2R} = h_l(CH_2THF)$$

$$r_3 = \frac{V_{MTHFR}}{(1+[1+\sum_1^5 \frac{x_n}{K_3^{MTX_n}} + \frac{DHF}{K_3^{DHF}}]K_{CH_2THF}/CH_2THF)(1+K_{NADPH}/NADPH)}$$

$$r_4 = \frac{V_{MTR}}{(1+K_{CH_2THF}/CH_2THF)(1+K_{HCYS}/Homocysteine)}$$

$$r_5 = \frac{V_{FTS}}{(1+K_{THF}/THF)(1+K_{ATP}/ATP)(1+K_{formate}/Formate)}$$

$$r_6 = \frac{V_{GART}}{\left(1+\left(1+\sum_1^5 \frac{x_n}{K_6^{MTX_n}} + \frac{DHF}{K_6^{DHF}} + \frac{CHODHF}{K_6^{CHODHF}}\right)(K_{CHOTHF}/CHOTHF)\right)(1+K_{GAR}/GAR)}$$

$$r_7 = \frac{V_{ATIC}}{\left(1+\left(1+\sum_1^5 \frac{x_n}{K_7^{MTX_n}}+\frac{DHF}{K_7^{DHF}}+\frac{CHODHF}{K_7^{CHODHF}}\right)(K_{CHOTHF}/CHOTHF)\right)(1+K_{AICAR}/AICAR)}$$

$$r_8 = \frac{V_{MTHFD}}{(1+K_{CH_2THF}/CH_2THF)(1+K_{NADP+}/NADP+)}$$

$$r_9 = \frac{V_{TS}}{\left(1+\sum_2^5 \frac{x_n}{K_9^{MTX_n}}+\frac{DHF}{K_9^{FH_2}}\right)\frac{K_{CH_2THF}K_{dUMP}}{dUMP\cdot CH_2THF}+\left(\left(1+\sum_2^5 \frac{x_n}{K_9^{MTX_n}}+\frac{DHF}{K_9^{DHF}}\right)\left(1+\frac{CHODHF}{K_9^{CHODHF}}\right)+\frac{CHODHF}{K_9^{CHODHF}}\cdot\frac{x_1}{K_9^{MTX_1}}\right)\frac{K_{CH_2THF}}{CH_2THF}+\left(1+\sum_1^5 \frac{x_n}{K_9^{MTX_n}}+\frac{DHF}{K_9^{DHF}}\right)}$$

$$r_{10} = k(DHF^T - DHF) \qquad r_{11} = k_{FDS}(DHF)$$

$$r_{12} = \frac{V_{ATIC12}}{\left(1+\left(1+\sum_1^5 \frac{x_n}{K_7^{MTX_n}}+\frac{DHF}{K_7^{DHF}}+\frac{CHOTHF}{K_{12}^{CHOTHF}}\right)(K_{CHODHF}/CHODHF)\right)(1+K_{AICAR}/AICAR)}$$

$$r_{13} = \frac{V_{FA}}{(1+K_{GLN}/GLN)(1+K_{FGAR}/FGAR)}.$$

where DHF^T-DHF (total DHF minus free DHF) is the concentration of DHF bound to DHFR, x_i is the concentration of *i*-glutamated MTX, and all folates are assumed to be penta-glutamated; not shown are 10 additional differential equations for up to penta-glutamation of MTX either free or bound to DHFR. Differences in the rational polynomial structures shown above imply differences in reaction mechanisms [1]. For steady state flux predictions, it is assumed that microarray averages correspond to the steady state of the model. For time course microarray data it is assumed that initial values in the array experiments correspond to the steady state of the model. Since the model has only one compartment, the cytosol, it cannot handle changes in the mitochondrial proteins MTHFD2 and SHMT2, nor can it handle changes in the extra-cellular folate hydrolase FOLH1 (the gene that codes for prostate-specific membrane antigen, PSMA), so these genes are ignored in the analyses. Flux boundary conditions for dUMP and GAR synthesis in the original model were replaced by downstream concentration boundary conditions set equal to their initial values because *steady state* flux differences would otherwise be nullified, e.g. if the flux into GAR were fixed, the steady state flux through GART would also be fixed, and artificially then, there would be no variation in predictions of this flux across patients.

Data Manipulations

MAS5 microarray measurements of Ross et al [7] and Yeoh et al [8] were normalized by dividing by the mean of medians within leukemia subtype. The resulting metrics were then used to modulate baseline folate model V_{max} values to create steady state folate predictions. The steady state predictions were computed as simulation endpoints at 40 hours post perturbation from the initial folate model operating point which was assumed to correspond to the mean of the medians of the leukemia groups. The folate genes GGH (polyglutamate hydrolase) , FPGS (polyglutamate synthase), and RFC (reduced folate carrier), were not considered because these reactions are included in the model only for MTX and not folate, and because MTX=0 in the microarray data analyses.

Systems Biology Markup Language for R

Systems biology markup language (SBML) is a standard for representing dynamical systems of biological interest [9-11]. The package SBMLR [12] provides an SBML-like R model structure, examples of simple analyses, and functions for exchanging models to and from SBML. The SBMLR package includes the folate metabolism model used here [1] and a model of purine metabolism [13]. R scripts used to produce figures 2-8 in this paper are available in the `BMCcancer04` directory of the SBMLR package. The R data packages used, and the latest version of SBMLR, can be obtained from the author's website [14].

Results

Folate system correlations across childhood leukemias

Childhood acute lymphoblastic leukemia (ALL) is treated with methotrexate. Thus, it seemed plausible to conjecture that differences in treatment outcome across leukemia

subtypes might be discernable as differences in folate gene expression patterns. To explore this, we investigated the childhood ALL microarray data of Ross et al [7] and Yeoh et al [8] shown in Figures 3 and 4; throughout this paper, genes with multiple probe sets will be represented by the set with the highest average value. Several points can be made regarding this steady state diagnostic bone marrow data. Firstly, since TYMS and DHFR (similarly MTHFD1, GART and ATIC) operate in series, it makes sense that the system will attempt to match these throughput capabilities as closely as possible to avoid the costs of maintaining unneeded excess "equipment." Thus, positive correlations *within* the DNPS and DNTS branches are expected. Secondly, growing cells require commensurate increases in both *de novo* dTMP and DNPS, so positive correlations *between* DNPS and DNTS are also expected. Finally, DNPS genes are higher in T cell leukemic cells than in B-cell leukemic cells, consistent with measured DNPS rates [15].

Since MTHFD1 and TYMS are the "gatekeepers" of DNPS and DNTS, respectively, correlation plots for these genes should be compared to corresponding model flux predictions. This is done in Figure 5 (see Methods for mathematical details) which shows that the model applied to the data in Figures 3 and 4 yields increased predicted DNPS vs. DNTS correlations relative to measured MTHFD1 vs. TYMS correlations. To estimate the amount of correlation attributable to steady state model flux constraints alone, 1000 uncorrelated simulated patients (normally distributed with standard deviations of 20%) were applied to the model in place of the datasets in Figures 3 and 4. The amount of correlation found (r=0.11, Figure 6) is consistent with model induced changes for BCR-ABL and T cell leukemias (Figure 5), but the correlations for TEL-AML1 leukemias increased by more than this amount, so for this leukemia type there must be additional information contributed by other folate genes inputted into the model.

Folate System Analysis of Radiation Time Course Data

To determine if folate correlations also exist in transient data, acute dose ionizing radiation response time course data [16] was investigated. The data (Fig. 7) shows that TS, DHFR, GARFT and MTHFD each have a dose-dependent transient increase after irradiation, consistent with radiation induced DNA damage causing a transient rise and fall in P53 activity [17] with subsequent induction of ribonucleotide reductase subunit P53R2 [18] and thus, presumably, increases in *de novo* deoxynucleotide synthesis for DNA repair; a 17-fold increase in R2 protein 24 h after IR has also been observed [19] and found to be radioprotective [20]. The data also shows a steady decline in many of the gene expression time courses, possibly due to radiation induced cell cycle arrest.

There is no biological rationale for qualitative differences between 3 and 10 gray gene expression time courses. Thus, gene expression time courses are more likely to be "real signals" if they differ between doses only in terms of minor time shifts and plausible (e.g. dose ordered) amplitude changes. Based on this criterion, MTHFR and SHMT1 (in Figure 7) were dismissed as noise. The remaining six gene expression time courses in Figure 7 were applied to Morrison's model as time-varying modulators of V_m values. The resulting plots are shown in Figure 8. These plots affirm the "dose-dependent spike resting on a decreasing backdrop" response of DNPS and DNTS that was qualitatively inferred from Figure 7.

Discussion

This paper showed how steady state childhood ALL patient data and transient in vitro radiation response data can be jointly mined through a common focus on folate system correlations. Pathway focused analyses are advantageous, if not critical, for gene-gene correlation studies because the number of possible correlations will otherwise become too great. For example, if a chip carries 10,000 genes, the number of 2D plots requiring

correlation testing is 10,000 choose 2, or ~50 million. The pathway specific analyses given here are exemplary feasible approaches to multivariate gene correlations studies.

Although TS and DHFR are predominantly controlled at the protein level, Figures 3 and 4 suggest that some control effort is also exerted on the mRNA level. Thus, even if protein level control dominates a particular regulatory system, mRNA signals can still be informative, possibly revealing what it is that the overall system is trying to accomplish. To carry this concept further, consider the following logic. The controllability of any process can only increase as the number of "valves" increases; one can always hold an extra valve fixed so it can never hurt to have it. Given that biochemical systems have "mRNA valves," such systems can only gain process controllability by using them. Thus, the odds are in favor of mRNA levels being used in control at least to some degree under most operating conditions. Control efforts implemented at the mRNA level may be merely "fine tuning adjustments," but such adjustments may nevertheless be informative signals of the control system's state. Based on this logic, DNA microarray data likely contains considerable amounts of untapped biochemical control system information. Multivariate analyses will be the key to extracting such information, and pathway focused approaches will enable such multivariate analyses.

Pathway focused analyses require some knowledge of the underlying biochemical literature, the more the better. To accrue such expertise, statistician microarray analysts should perhaps heed the control system engineer's adage that "you must marry a process." If this can be accomplished across the field, "omic" data analyses of the future could then be carried out using a "quilt" of pathway focused experts. Once established, all publicly available DNA microarray data could be processed by this "quilt." Convergence toward this scenario seems

inevitable, since the current approach, in which statisticians are expected to make sense of all of the genes all at once, is far too overwhelming, and indeed, unrealistic.

As food for additional speculative thought, if one considers that BCR-ABL (non-curable) and TEL-AML1 (curable) leukemias are both pro B cell leukemias, and thus that differentiation state changes would not be required in attempts to transfer poor prognosis BCR-ABL cell states toward TEL-AML1 states, a treatment-failure risk state transfer therapeutic strategy can be envisioned whereby diagnostic "initial state" microarray measurements are used to suggest patient specific multi-drug dose timing schedules that would possibly drive BCR-ABL patient states toward the TEL-AML1 population average state. Standard therapy could then be applied, presumably with increased chances of success. From a modeling perspective, the elegance of this proposed approach is that the "solved" TEL-AML1 counterpart to the BCR-ABL problem provides a target state *within* the modeled system (and thus a minimal distance objective function in the context of optimal control theory). For most cancers, solved counterparts do not exist, and the challenge in this general case lies in the demand for models which entail everything between the drug target and cell death, and with such accuracy that differences can be predicted between malignant and normal cells. Currently, this is not feasible. Although the scarcity of cancers with solved counterparts could perhaps be surmounted by fractionating patients as curable and non-curable using microarrays [21-23] rather than chromosomal translocations, a fundamental flaw with this proposed framework is the underlying assumption that differences in treatment outcome will lie within the biochemical system being targeted by the therapy. For example, with respect to folate flux predictions, differences in outcome may lie downstream of DNA damage and repair and thus outside of the folate model. Evidence for this in the Ross et al data (not shown) lies in observations that hyperploidy (curable) leukemia flux states fall over BCR-ABL states rather

than TEL-AML1 states, and that TEL-AML1 failed patients are not closer to the BCR-ABL center, and nor is the one BCR-ABL cured patient closer to the TEL-AML1 center. The risk-state-transfer framework is, however, newly sprouted, so it should be given a chance to grow, particularly since models will inevitably expand to include more and more of their neighboring processes; for example, the folate model will eventually become part of a folate-purine-pyrimidine model with a focus on deoxynucleotides, which will eventually include DNA damage and repair, which will eventually include extensions to apoptosis, and so on, until it eventually encompasses systems which do contain treatment failure differences. Folate models can be viewed as the "front-end" to any modeling of MTX responses. One could logically argue that a modeling sweep that begins at the beginning is a sound approach.

Conclusions

The main conclusion of this paper is that interesting inferences can be gleaned from genome-wide microarray data (with or without mathematical models) if gene-gene correlations are analyzed in a pathway specific manner. The added value of analyzing microarray data using Morrison's folate model, relative to simply "eyeballing" the gene expression data, was minimal. In figure 5, gate-keeper focused gene expression scatter plots are basically as revealing as model-predicted DNPS vs. DNTS scatter plots. Similarly, for the radiation time course data in Figure 7, the spike increase in DNPS and DNTS and the baseline downward trend at larger times are both also apparent in the data. Thus, the value of pathway specific focused analyses of DNA microarray data lies, to a large extent, in focused multivariate analyses of the data itself. Nevertheless, to go beyond the qualitative statements made above, and to actually plot predicted fluxes (Figures 5 and 8), a quantitative model is needed. Further, as models continue to expand in their scope and complexity, *gedanken* experiments underlying "eye-ball" data analyses will become more and more difficult to carry out.

Mathematical models will then become indispensable. In this context, this paper can be viewed as a positive control experiment whereby modeling yields expected results.

Competing interests

TR has no competing interests

Authors' contributions

TR is the sole contributor.

Acknowledgements

This research was supported by the Biostatistics Core Facility of the Comprehensive Cancer Center of Case Western Reserve University and University Hospitals of Cleveland (P30 CA43703), by the American Cancer Society (IRG-91-022-09), and by the National Cancer Institute's Integrative Cancer Biology Program (P20 CA112963-01).

References

1. Morrison PF, Allegra CJ: **Folate cycle kinetics in human breast cancer cells**. *JBiolChem* 1989, **264**(18):10552-10566.
2. Curtin NJ, Hughes AN: **Pemetrexed disodium, a novel antifolate with multiple targets**. *Lancet Oncol* 2001, **2**(5):298-306.
3. Shih C, Habeck LL, Mendelsohn LG, Chen VJ, Schultz RM: **Multiple folate enzyme inhibition: mechanism of a novel pyrrolopyrimidine-based antifolate LY231514 (MTA)**. *Adv Enzyme Regul* 1998, **38**:135-152.
4. Yin MB, Guimaraes MA, Zhang ZG, Arredondo MA, Rustum YM: **Time dependence of DNA lesions and growth inhibition by ICI D1694, a new quinazoline antifolate thymidylate synthase inhibitor**. *Cancer Res* 1992, **52**(21):5900-5905.
5. Spiegelman S, Sawyer R, Nayak R, Ritzi E, Stolfi R, Martin D: **Improving the anti-tumor activity of 5-fluorouracil by increasing its incorporation into RNA via metabolic modulation**. *ProcNatlAcadSciUSA* 1980, **77**(8):4966-4970.
6. Radivoyevitch T: **Sphingoid base metabolism in yeast: Mapping gene expression patterns into qualitative metabolite time course predictions**. *Comparative & Functional Genomics* 2001, **2**:289-294.
7. Ross ME, Zhou X, Song G, Shurtleff SA, Girtman K, Williams WK, Liu HC, Mahfouz R, Raimondi SC, Lenny N, Patel A, Downing JR: **Classification of pediatric acute lymphoblastic leukemia by gene expression profiling**. *Blood* 2003, **102**(8):2951-2959.
8. Yeoh EJ, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, Behm FG, Raimondi SC, Relling MV, Patel A, Cheng C, Campana D, Wilkins D, Zhou X, Li J, Liu H, Pui CH, Evans WE, Naeve C, Wong L, Downing JR: **Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling**. *Cancer Cell* 2002, **1**(2):133-143.
9. **Systems Biology Markup Language** [http://sbml.org]
10. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin, II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novere N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J: **The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models**. *Bioinformatics* 2003, **19**(4):524-531.
11. Finney A, Hucka M: **Systems biology markup language: Level 2 and beyond**. *Biochem Soc Trans* 2003, **31**(Pt 6):1472-1473.
12. **SBMLR** [http://www.bioconductor.org/repository/devel/package/html/SBMLR.html]
13. Curto R, Voit EO, Sorribas A, Cascante M: **Validation and steady-state analysis of a power-law model of purine metabolism in man**. *BiochemJ* 1997, **324 (Pt 3)**:761-775.
14. **Radivoyevitch Lab** [http://epbi-radivot.cwru.edu/]
15. Dervieux T, Brenner TL, Hon YY, Zhou Y, Hancock ML, Sandlund JT, Rivera GK, Ribeiro RC, Boyett JM, Pui CH, Relling MV, Evans WE: **De novo purine synthesis inhibition and antileukemic effects of mercaptopurine alone or in combination with methotrexate in vivo**. *Blood* 2002, **100**(4):1240-1247.

16. Jen KY, Cheung VG: **Transcriptional response of lymphoblastoid cells to ionizing radiation**. *Genome Res* 2003, **13**(9):2092-2100.
17. Lahav G, Rosenfeld N, Sigal A, Geva-Zatorsky N, Levine AJ, Elowitz MB, Alon U: **Dynamics of the p53-Mdm2 feedback loop in individual cells**. *Nat Genet* 2004, **36**(2):147-150.
18. Tanaka H, Arakawa H, Yamaguchi T, Shiraishi K, Fukuda S, Matsui K, Takei Y, Nakamura Y: **A ribonucleotide reductase gene involved in a p53-dependent cell-cycle checkpoint for DNA damage**. *Nature* 2000, **404**(6773):42-49.
19. Kuo ML, Kinsella TJ: **Expression of ribonucleotide reductase after ionizing radiation in human cervical carcinoma cells**. *Cancer Res* 1998, **58**(10):2245-2252.
20. Kuo ML, Hwang HS, Sosnay PR, Kunugi KA, Kinsella TJ: **Overexpression of the R2 subunit of ribonucleotide reductase in human nasopharyngeal cancer cells reduces radiosensitivity**. *Cancer J* 2003, **9**(4):277-285.
21. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM: **A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma**. *Proc Natl Acad Sci U S A* 2003, **100**(17):9991-9996.
22. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltnane JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, Lopez-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T, Staudt LM: **The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma**. *N Engl J Med* 2002, **346**(25):1937-1947.
23. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Jr., Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Staudt LM, . **Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling**. *Nature* 2000, **403**(6769):503-511.

Figures

Figure 1. The folate cycle model of Morrison and Allegra [1] (A) and the molecular structure of folate (B).

Figure 2. Morrison and Allegra's model [1] responding to 1 μM MTX applied continuously after t=0. Concentrations are in μM (top 6 plots) and fluxes are in μM/hour (bottom 3 plots).

Figure 3. MAS5 U133a folate gene expression data of Ross et al [7]. Symbols are TEL-AML1 (B), BCR-ABL (b) and T-cell (T).

Figure 4. MAS5 U95av2 folate gene expression data of Yeoh et al [8]. Only the Yeoh et al [8] patients who are also in the Ross et al dataset [7] are considered. MAS5 summary measures were computed from cel files using Bioconductor's AFFY package. Symbols are as in Figure 3.

Figure 5. Comparisons of measured MTHFD1 vs. TYMS versus predicted DNPS vs. DNTS.

Figure 6. DNPS vs. DNTS predicted with randomly driven genes replacing Figures 3 or 4.

Figure 7. Lymphocyte radiation response data of Jen and Cheung (2003). SHMT1 and MTHFR were not applied to the folate model, see text.

Figure 8. The folate model's response to the radiation time course data. Fluxes are in μM/hr.

Additional files

The following zip files are available through the author's website [14] for analysts interested in reproducing the results.

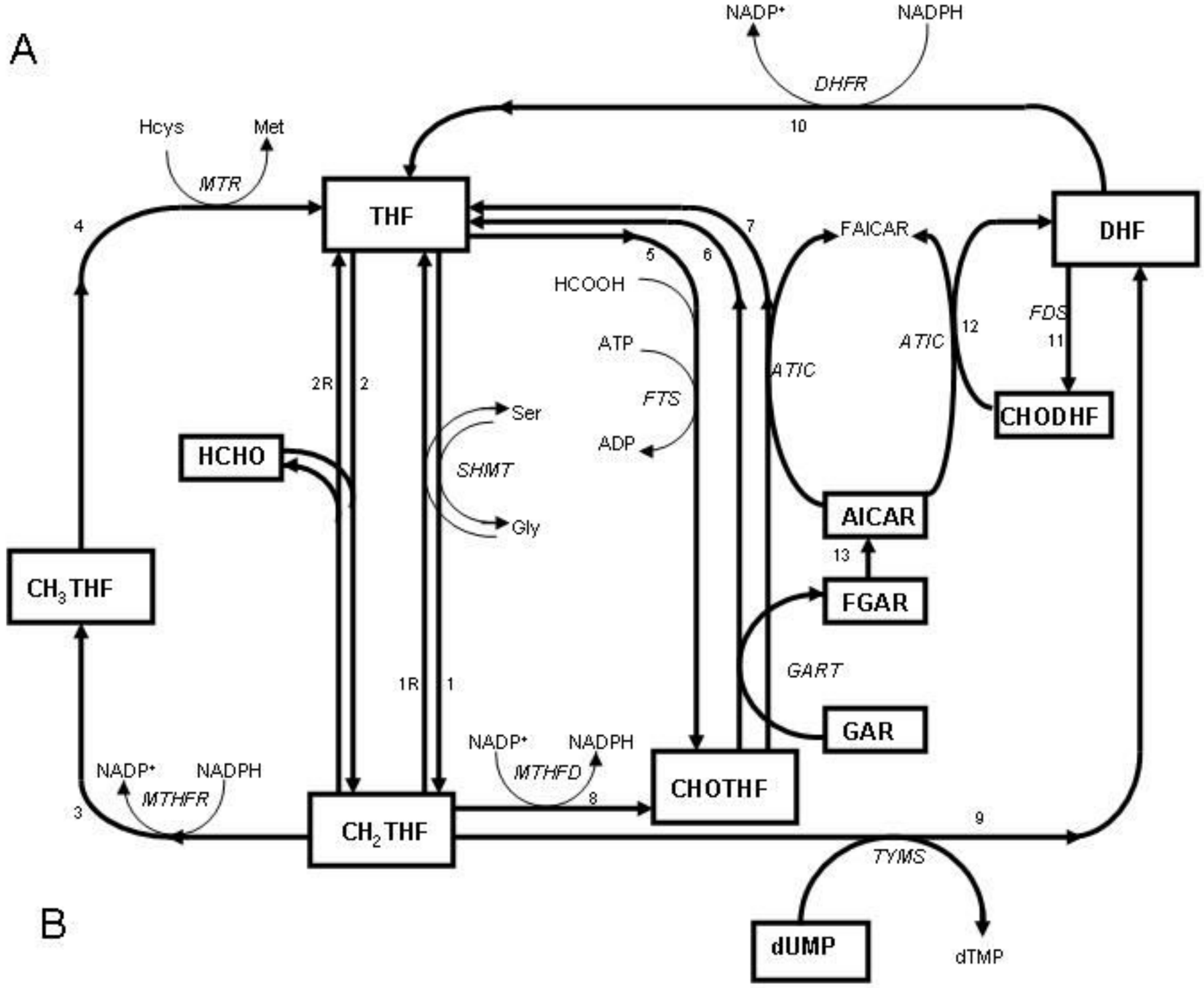
SBMLR.zip
R package SBMLR, version 1.12. This requires the XML and ODESOLVE packages. R scripts used to generate figures 2 through 8 are now included (along with this manuscript) in the `BMCcancer04` subdirectory of the SBMLR package installation.

cheungEset.zip
R data package for Jen and Cheung's radiation time course data.

rossEset.zip
R data package for Ross et al's leukemia data.

yeohEset.zip
R data package for Yeoh et al's leukemia data.

A


NADP+
NADPH
DHFR
10
Hcys
Met
MTR
4
THF
7
FAICAR
DHF
5
6
HCOOH
ATP
FTS
ADP
ATIC
ATIC
12
FDS
11
2R
2
Ser
SHMT
Gly
HCHO
CHODHF
AICAR
13
FGAR
CH3THF
GART
GAR
1R
1
NADP+
NADPH
MTHFD
8
NADP+
NADPH
MTHFR
3
CHOTHF
9
CH2THF
TYMS
dUMP
dTMP

B

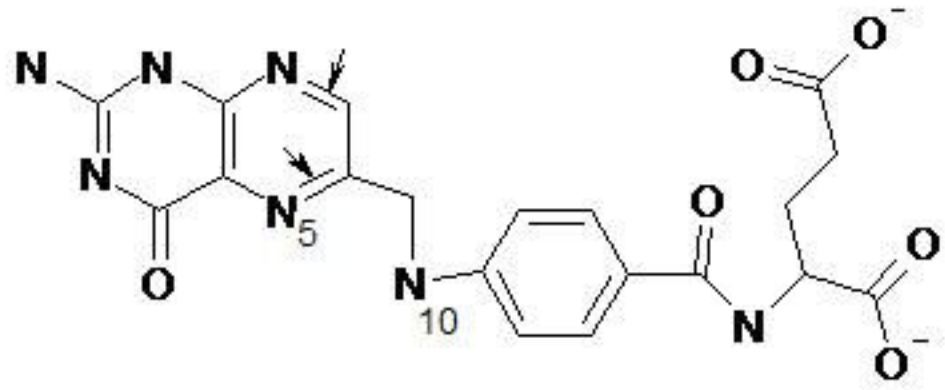
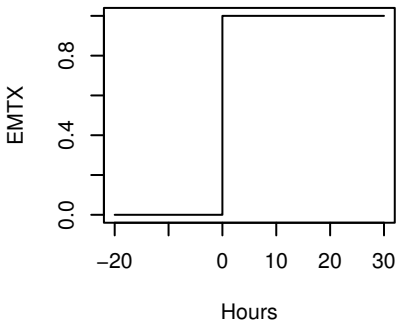
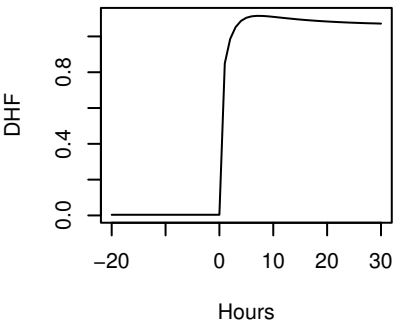
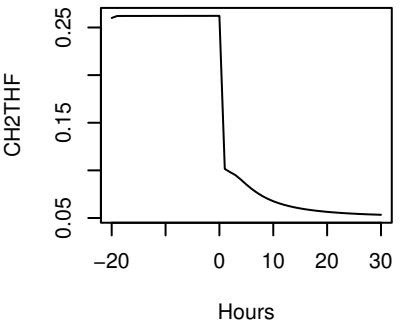
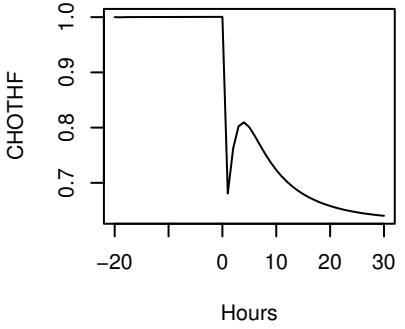

N
N
N
N5
O
N
10
O
O
N
O
O
O
O
-

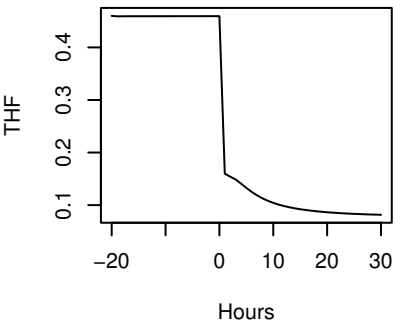
Figure 1

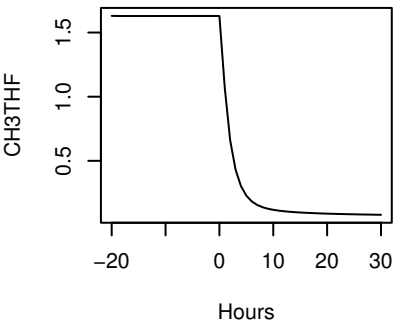

EMTX
0.0
0.4
0.8
−20
0
10
20
30
Hours

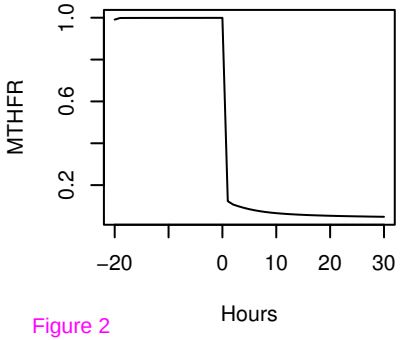

DHF
0.0
0.4
0.8
−20
0
10
20
30
Hours

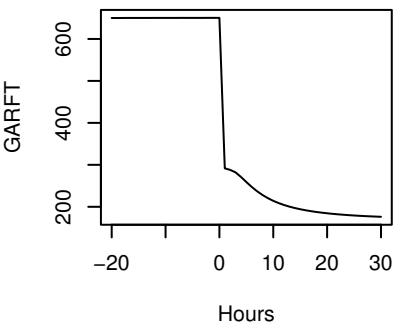

CH2THF
0.05
0.15
0.25
−20
0
10
20
30
Hours


CHOTHF
0.7
0.8
0.9
1.0
−20
0
10
20
30
Hours


THF
0.1
0.2
0.3
0.4
−20
0
10
20
30
Hours


CH3THF
0.5
1.0
1.5
−20
0
10
20
30
Hours


MTHFR
0.2
0.6
1.0
−20
0
10
20
30
Hours


GARFT
200
400
600
−20
0
10
20
30
Hours

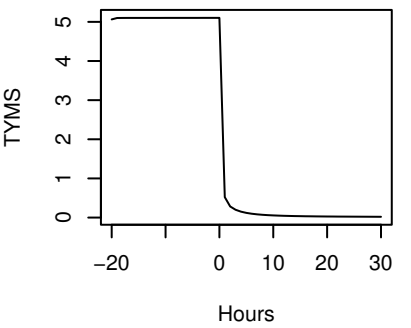

TYMS
0
1
2
3
4
5
−20
0
10
20
30
Hours

Figure 2

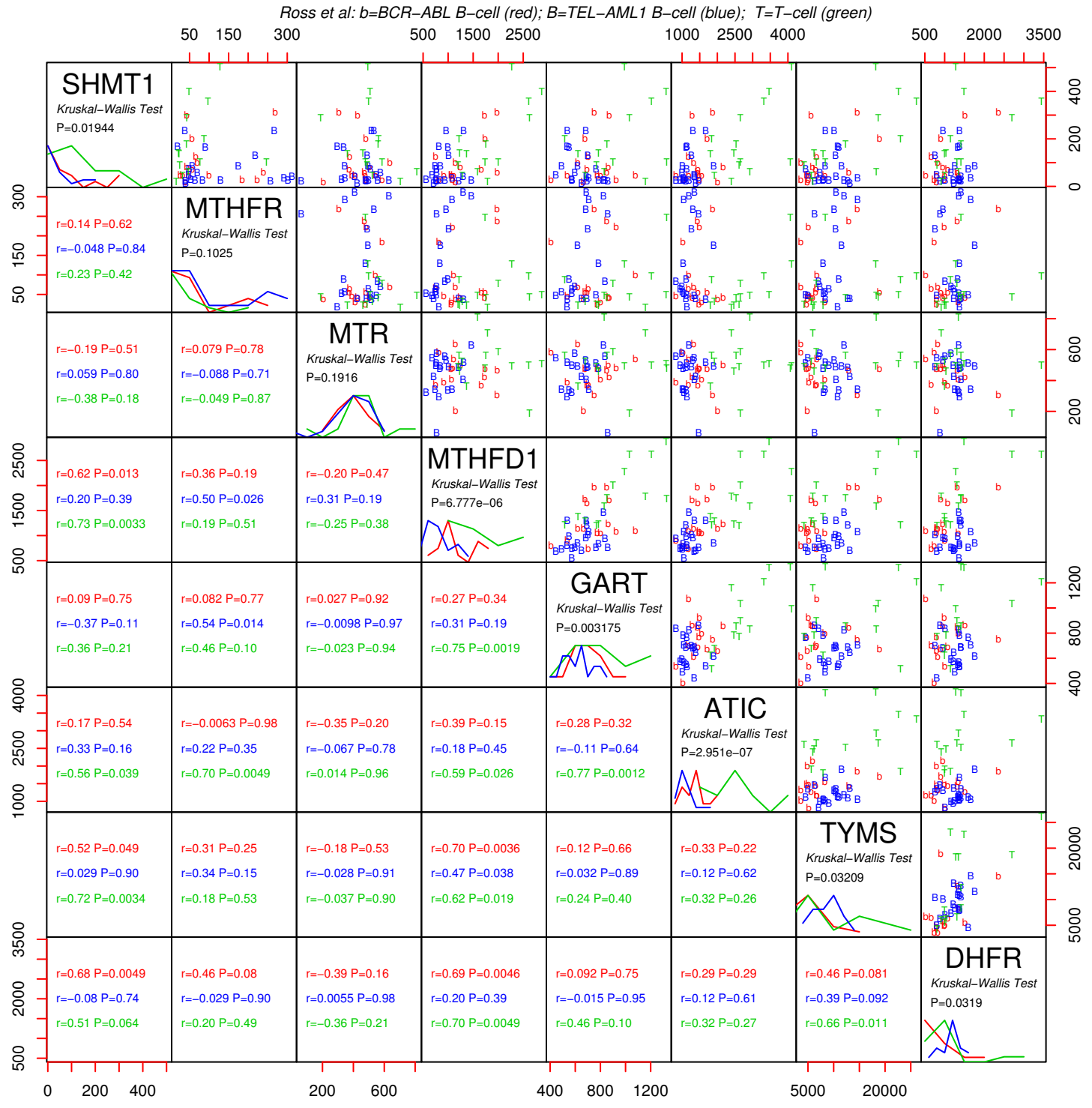

Ross et al: b=BCR−ABL B−cell (red); B=TEL−AML1 B−cell (blue); T=T−cell (green)
SHMT1
Kruskal−Wallis Test
P=0.01944
MTHFR
Kruskal−Wallis Test
P=0.1025
MTR
Kruskal−Wallis Test
P=0.1916
MTHFD1
Kruskal−Wallis Test
P=6.777e−06
GART
Kruskal−Wallis Test
P=0.003175
ATIC
Kruskal−Wallis Test
P=2.951e−07
TYMS
Kruskal−Wallis Test
P=0.03209
DHFR
Kruskal−Wallis Test
P=0.0319
r=0.14 P=0.62
r=−0.048 P=0.84
r=0.23 P=0.42
r=−0.19 P=0.51
r=0.059 P=0.80
r=−0.38 P=0.18
r=0.079 P=0.78
r=−0.088 P=0.71
r=−0.049 P=0.87
r=0.62 P=0.013
r=0.20 P=0.39
r=0.73 P=0.0033
r=0.36 P=0.19
r=0.50 P=0.026
r=0.19 P=0.51
r=−0.20 P=0.47
r=0.31 P=0.19
r=−0.25 P=0.38
r=0.09 P=0.75
r=−0.37 P=0.11
r=0.36 P=0.21
r=0.082 P=0.77
r=0.54 P=0.014
r=0.46 P=0.10
r=0.027 P=0.92
r=−0.0098 P=0.97
r=−0.023 P=0.94
r=0.27 P=0.34
r=0.31 P=0.19
r=0.75 P=0.0019
r=0.17 P=0.54
r=0.33 P=0.16
r=0.56 P=0.039
r=−0.0063 P=0.98
r=0.22 P=0.35
r=0.70 P=0.0049
r=−0.35 P=0.20
r=−0.067 P=0.78
r=0.014 P=0.96
r=0.39 P=0.15
r=0.18 P=0.45
r=0.59 P=0.026
r=0.28 P=0.32
r=−0.11 P=0.64
r=0.77 P=0.0012
r=0.52 P=0.049
r=0.029 P=0.90
r=0.72 P=0.0034
r=0.31 P=0.25
r=0.34 P=0.15
r=0.18 P=0.53
r=−0.18 P=0.53
r=−0.028 P=0.91
r=−0.037 P=0.90
r=0.70 P=0.0036
r=0.47 P=0.038
r=0.62 P=0.019
r=0.12 P=0.66
r=0.032 P=0.89
r=0.24 P=0.40
r=0.33 P=0.22
r=0.12 P=0.62
r=0.32 P=0.26
r=0.68 P=0.0049
r=−0.08 P=0.74
r=0.51 P=0.064
r=0.46 P=0.08
r=−0.029 P=0.90
r=0.20 P=0.49
r=−0.39 P=0.16
r=0.0055 P=0.98
r=−0.36 P=0.21
r=0.69 P=0.0046
r=0.20 P=0.39
r=0.70 P=0.0049
r=0.092 P=0.75
r=−0.015 P=0.95
r=0.46 P=0.10
r=0.29 P=0.29
r=0.12 P=0.61
r=0.32 P=0.27
r=0.46 P=0.081
r=0.39 P=0.092
r=0.66 P=0.011
50 150 300
500 1500 2500
1000 2500 4000
500 2000 3500
0 200 400
50 150 300
200 600
200 400 600
500 1500 2500
400 800 1200
1000 2500 4000
400 800 1200
5000 20000
500 2000 3500
0 5000 20000

Figure 3

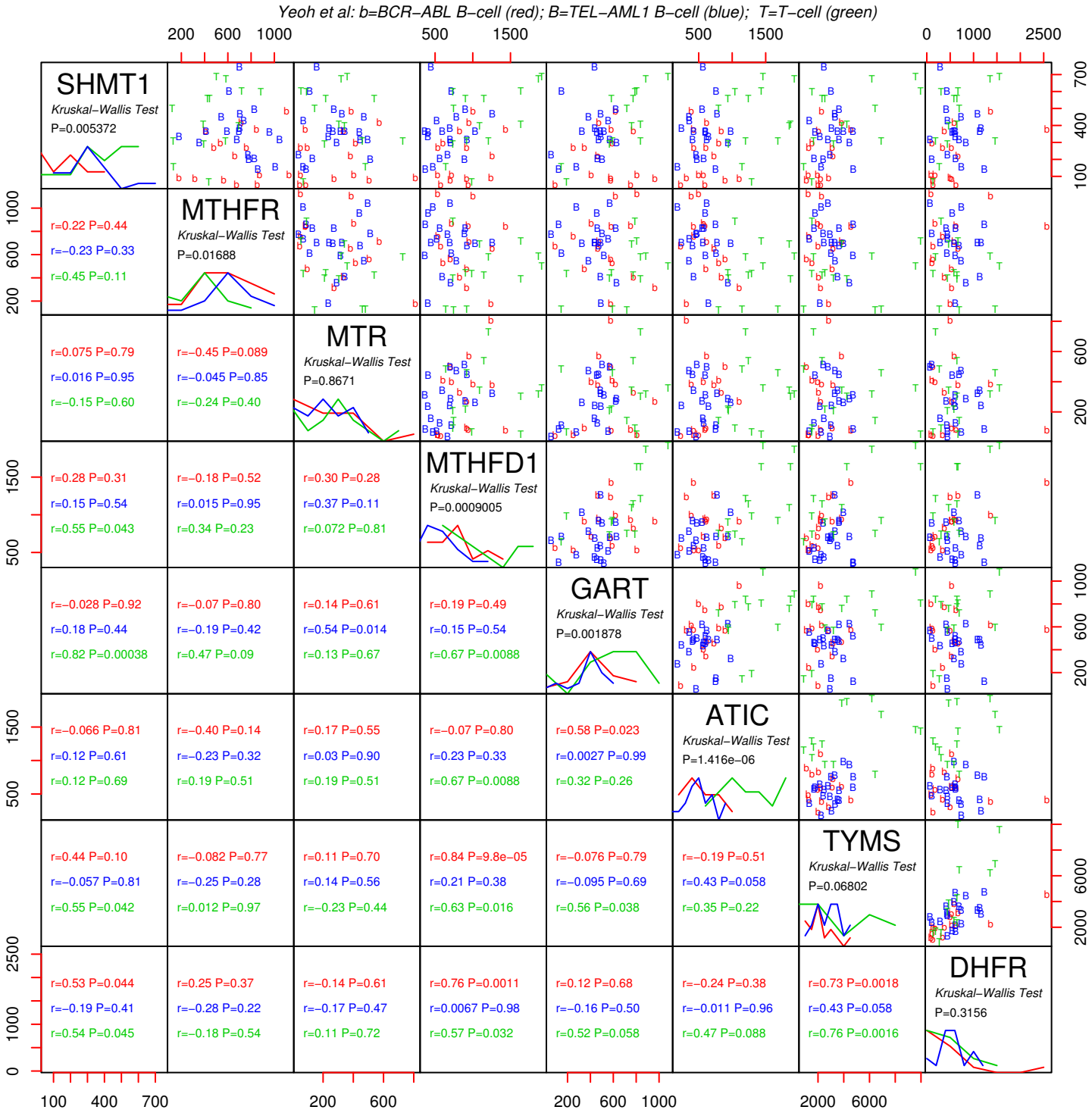
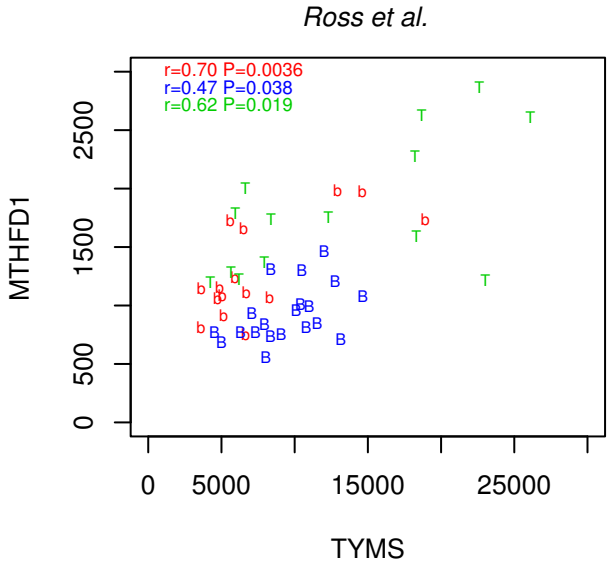
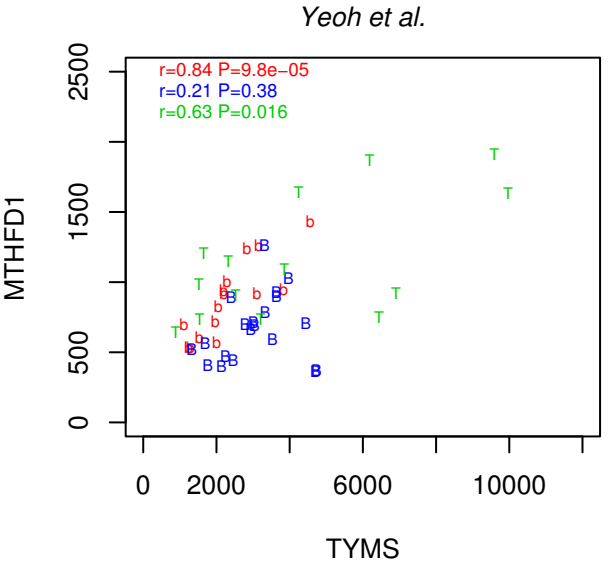
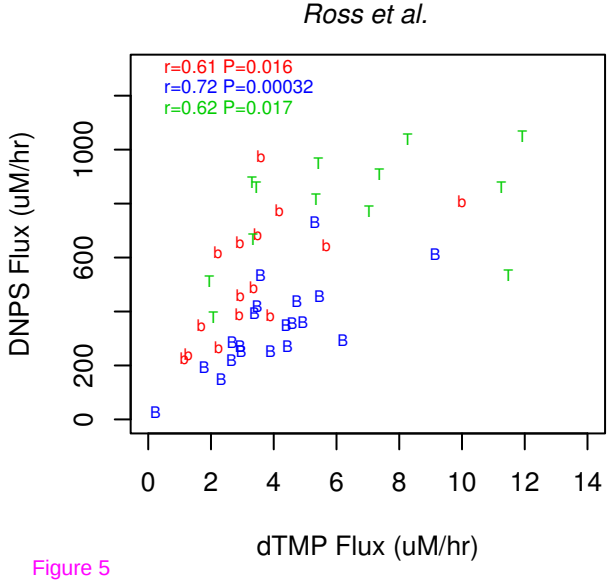

Yeoh et al: b=BCR−ABL B−cell (red); B=TEL−AML1 B−cell (blue); T=T−cell (green)
SHMT1
Kruskal−Wallis Test
P=0.005372
MTHFR
Kruskal−Wallis Test
P=0.01688
MTR
Kruskal−Wallis Test
P=0.8671
MTHFD1
Kruskal−Wallis Test
P=0.0009005
GART
Kruskal−Wallis Test
P=0.001878
ATIC
Kruskal−Wallis Test
P=1.416e−06
TYMS
Kruskal−Wallis Test
P=0.06802
DHFR
Kruskal−Wallis Test
P=0.3156
r=0.22 P=0.44
r=−0.23 P=0.33
r=0.45 P=0.11
r=0.075 P=0.79
r=0.016 P=0.95
r=−0.15 P=0.60
r=−0.45 P=0.089
r=−0.045 P=0.85
r=−0.24 P=0.40
r=0.28 P=0.31
r=0.15 P=0.54
r=0.55 P=0.043
r=−0.18 P=0.52
r=0.015 P=0.95
r=0.34 P=0.23
r=0.30 P=0.28
r=0.37 P=0.11
r=0.072 P=0.81
r=−0.028 P=0.92
r=0.18 P=0.44
r=0.82 P=0.00038
r=−0.07 P=0.80
r=−0.19 P=0.42
r=0.47 P=0.09
r=0.14 P=0.61
r=0.54 P=0.014
r=0.13 P=0.67
r=0.19 P=0.49
r=0.15 P=0.54
r=0.67 P=0.0088
r=−0.066 P=0.81
r=0.12 P=0.61
r=0.12 P=0.69
r=−0.40 P=0.14
r=−0.23 P=0.32
r=0.19 P=0.51
r=0.17 P=0.55
r=0.03 P=0.90
r=0.19 P=0.51
r=−0.07 P=0.80
r=0.23 P=0.33
r=0.67 P=0.0088
r=0.58 P=0.023
r=0.0027 P=0.99
r=0.32 P=0.26
r=0.44 P=0.10
r=−0.057 P=0.81
r=0.55 P=0.042
r=−0.082 P=0.77
r=−0.25 P=0.28
r=0.012 P=0.97
r=0.11 P=0.70
r=0.14 P=0.56
r=−0.23 P=0.44
r=0.84 P=9.8e−05
r=0.21 P=0.38
r=0.63 P=0.016
r=−0.076 P=0.79
r=−0.095 P=0.69
r=0.56 P=0.038
r=−0.19 P=0.51
r=0.43 P=0.058
r=0.35 P=0.22
r=0.53 P=0.044
r=−0.19 P=0.41
r=0.54 P=0.045
r=0.25 P=0.37
r=−0.28 P=0.22
r=−0.18 P=0.54
r=−0.14 P=0.61
r=−0.17 P=0.47
r=0.11 P=0.72
r=0.76 P=0.0011
r=0.0067 P=0.98
r=0.57 P=0.032
r=0.12 P=0.68
r=−0.16 P=0.50
r=0.52 P=0.058
r=−0.24 P=0.38
r=−0.011 P=0.96
r=0.47 P=0.088
r=0.73 P=0.0018
r=0.43 P=0.058
r=0.76 P=0.0016
200 600 1000
500 1500
500 1500
0 1000 2500
100 400 700
200 600
200 600 1000
2000 6000

Figure 4


Ross et al.
r=0.70 P=0.0036
r=0.47 P=0.038
r=0.62 P=0.019
MTHFD1
0
500
1500
2500
0
5000
15000
25000
TYMS


Yeoh et al.
r=0.84 P=9.8e−05
r=0.21 P=0.38
r=0.63 P=0.016
MTHFD1
0
500
1500
2500
0
2000
6000
10000
TYMS


Ross et al.
r=0.61 P=0.016
r=0.72 P=0.00032
r=0.62 P=0.017
DNPS Flux (uM/hr)
0
200
600
1000
0
2
4
6
8
10
12
14
dTMP Flux (uM/hr)

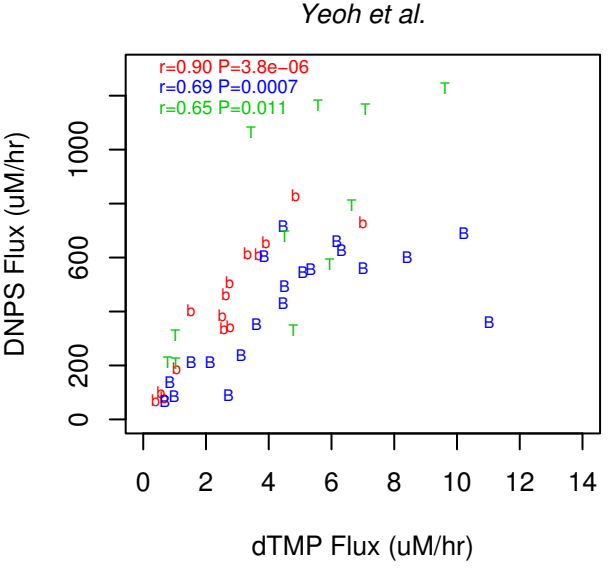

Yeoh et al.
r=0.90 P=3.8e−06
r=0.69 P=0.0007
r=0.65 P=0.011
DNPS Flux (uM/hr)
0
200
600
1000
0
2
4
6
8
10
12
14
dTMP Flux (uM/hr)

Figure 5

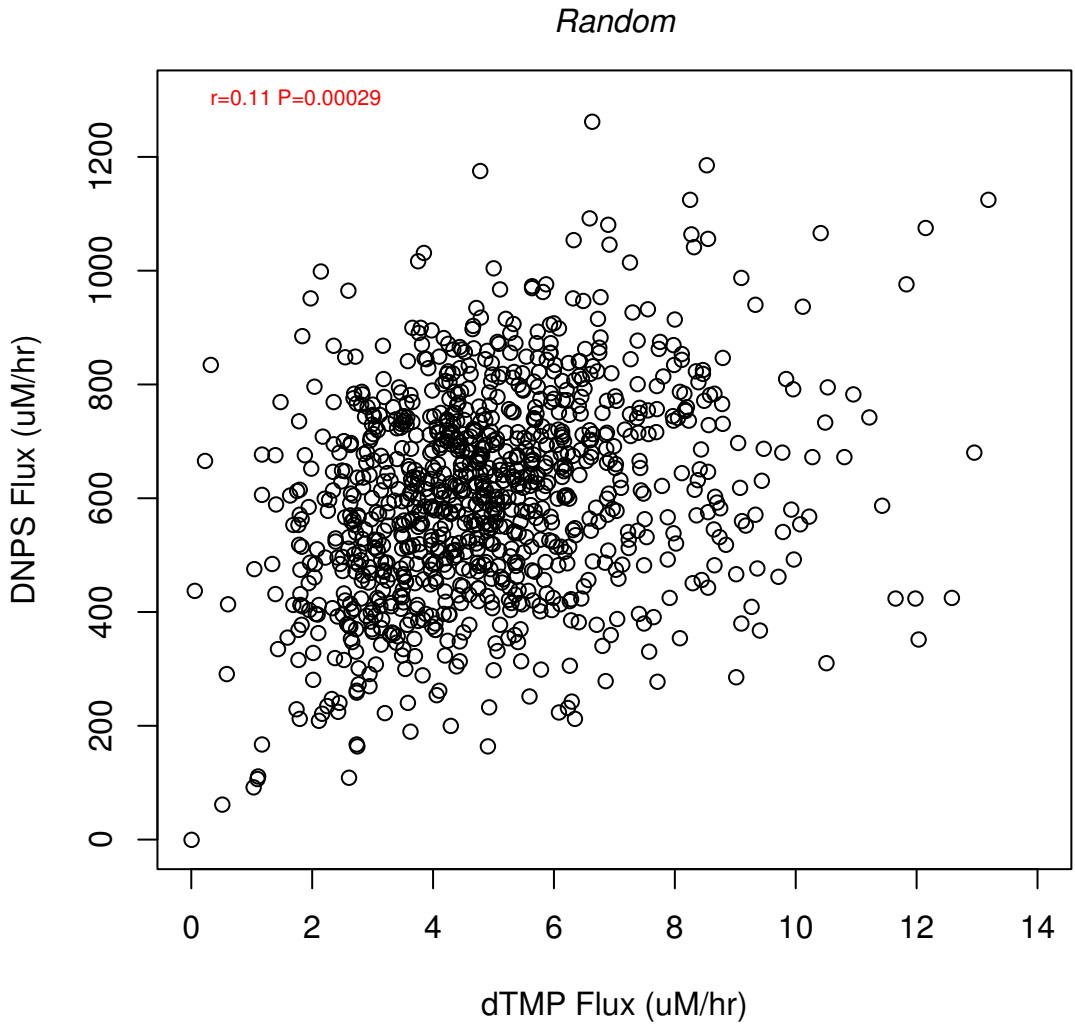
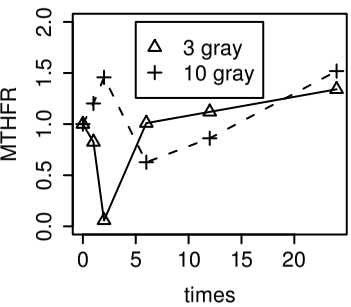
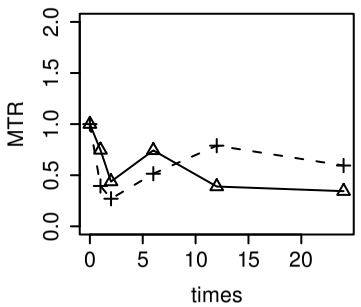
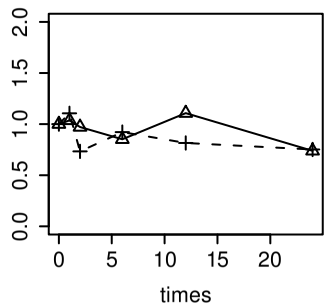
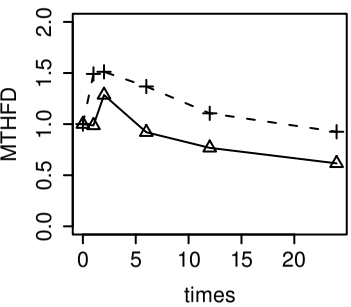

Random
r=0.11 P=0.00029
DNPS Flux (uM/hr)
0
200
400
600
800
1000
1200
0
2
4
6
8
10
12
14
dTMP Flux (uM/hr)

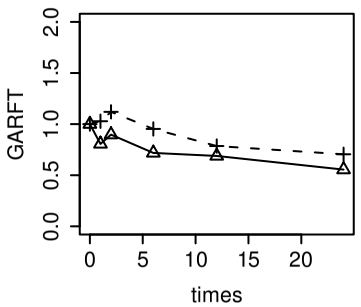
Figure 6

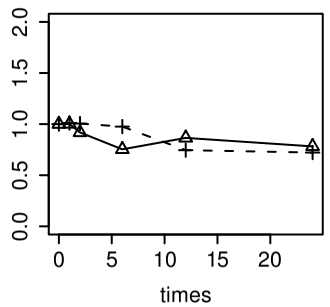

MTHFR
3 gray
10 gray
0.0
0.5
1.0
1.5
2.0
0
5
10
15
20
times

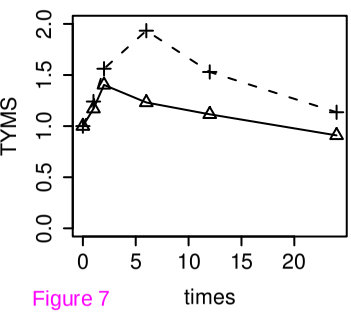

MTR
0.0
0.5
1.0
1.5
2.0
0
5
10
15
20
times


SHMT
0.0
0.5
1.0
1.5
2.0
0
5
10
15
20
times


MTHFD
0.0
0.5
1.0
1.5
2.0
0
5
10
15
20
times


GARFT
0.0
0.5
1.0
1.5
2.0
0
5
10
15
20
times


ATIC7
0.0
0.5
1.0
1.5
2.0
0
5
10
15
20
times


TYMS
0.0
0.5
1.0
1.5
2.0
0
5
10
15
20
times

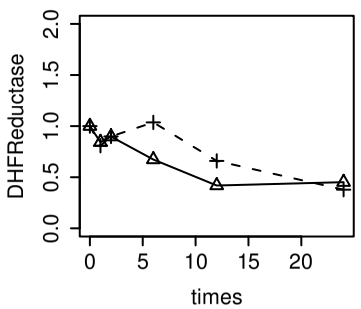
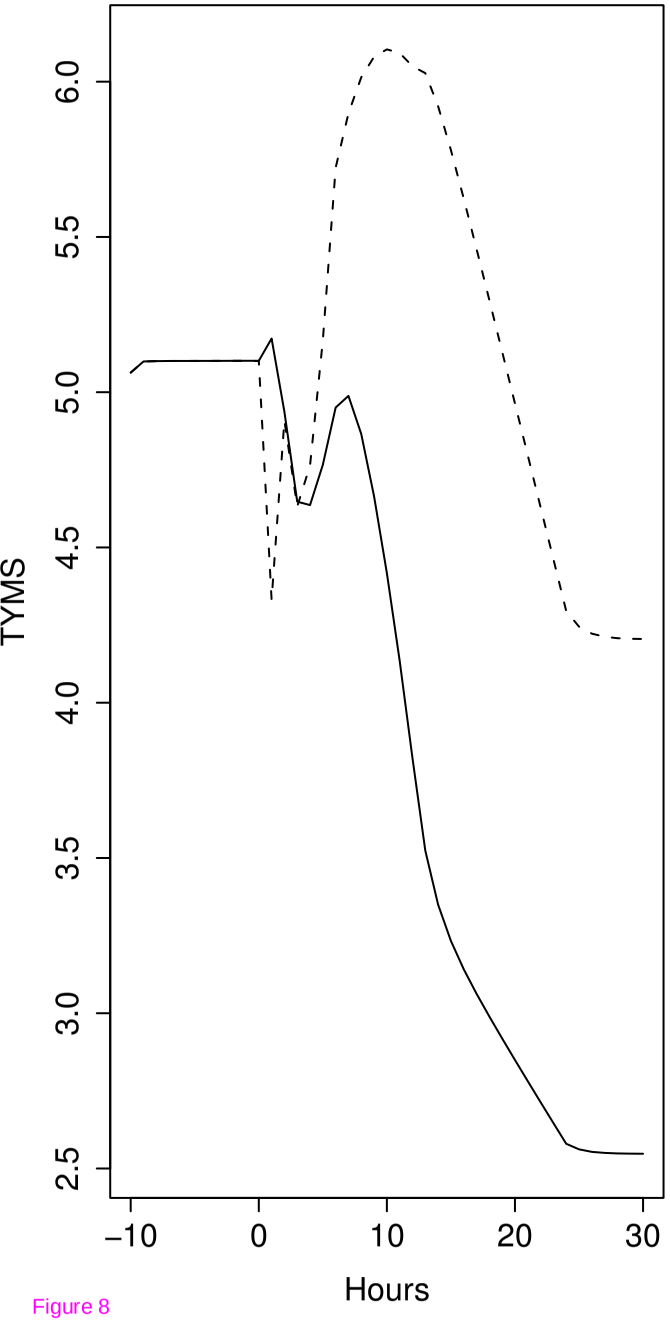

DHFReductase
0.0
0.5
1.0
1.5
2.0
0
5
10
15
20
times

Figure 7


TYMS
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
−10
0
10
20
30
Hours

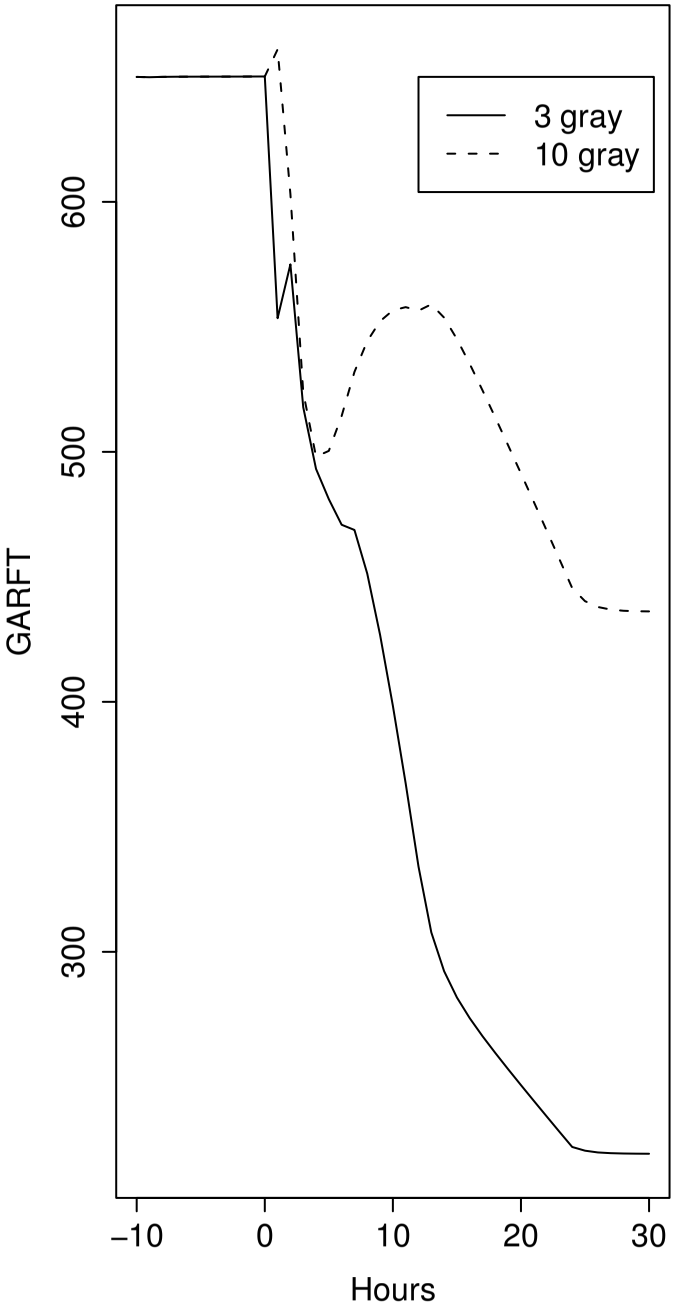

GARFT
600
500
400
300
3 gray
10 gray
−10
0
10
20
30
Hours

Figure 8