

Genetikai adatelemzések alkalmazásszintű párhuzamosítással

Dr. Antal Péter,
Hajós Gergely

Budapesti Műszaki és Gazdaságtudományi Egyetem
Méréstechnika és Információs Rendszerek Tanszék

GENAGRID

<http://genagrid.eu>

Bevezetés



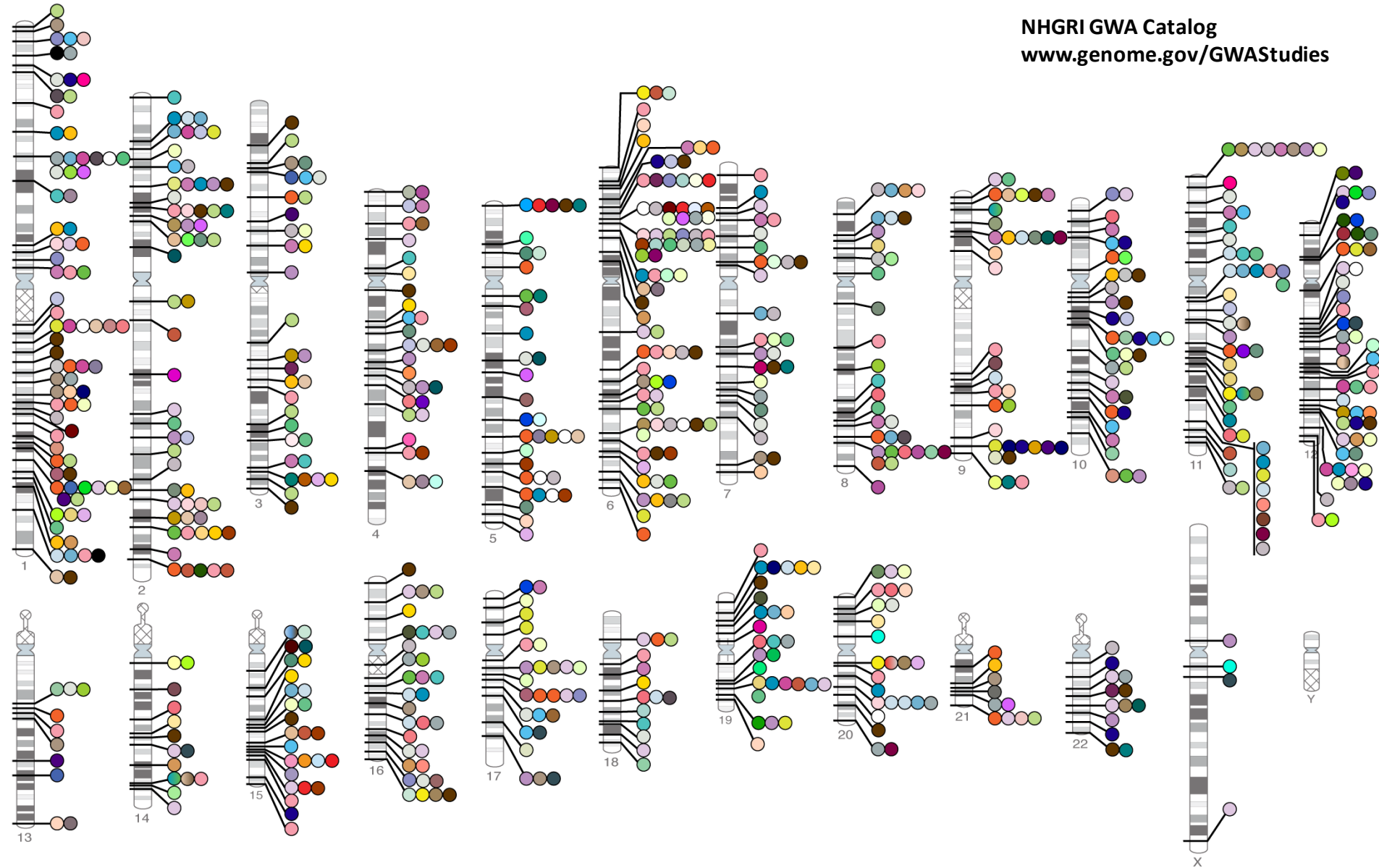
- Omikai kísérlettervezési és adatelemzési feladatok
 - Viszonylag kevés adat
 - Asszociációs vizsgálatok
 - Bayes-hálók
 - Alkalmazás szintű párhuzamosítás
- Hardverkörnyezet
 - SGI Altix ICE
 - 512 mag, 1 TB memória, 6.5 TFlops
- Szoftverkörnyezet
 - Condor, Torque (PBS)
- Terhelés
 - Egy számítás: 20-200 job
 - Egy job: 12-200 óra

Genome-wide associations studies 2010

779 studies, $p \leq 5 \times 10^{-8}$ 148 markers

NHGRI GWA Catalog

www.genome.gov/GWASudies



- Acute lymphoblastic leukemia
- Adhesion molecules
- Adiponectin levels
- Age-related macular degeneration
- AIDS progression
- Alcohol dependence
- Alzheimer disease
- Amyotrophic lateral sclerosis
- Angiotensin-converting enzyme activity
- Ankylosing spondylitis
- Arterial stiffness
- Asthma
- Atherosclerosis in HIV
- Atrial fibrillation
- Attention deficit hyperactivity disorder
- Autism
- Basal cell cancer
- Bipolar disorder
- Bilirubin
- Bladder cancer
- Blond or brown hair
- Blood pressure
- Blue or green eyes
- BMI, waist circumference
- Bone density
- Breast cancer
- C-reactive protein
- Cardiac structure/function
- Carnitine levels
- Carotenoid/tocopherol levels
- Celiac disease
- Chronic lymphocytic leukemia
- Cleft lip/palate
- Cognitive function
- Colorectal cancer
- Coronary disease
- Creutzfeldt-Jakob disease
- Crohn's disease
- Cutaneous nevi
- Dermatitis
- Drug-induced liver injury
- Eosinophil count
- Eosinophilic esophagitis
- Erythrocyte parameters
- Esophageal cancer
- Essential tremor
- Exfoliation glaucoma
- F cell distribution
- Fibrinogen levels
- Foliate pathway vitamins
- Freckles and burning
- Gallstones
- Glioma
- Glycemic traits
- Hair color
- Hair morphology
- HDL cholesterol
- Heart rate
- Height
- Hemostasis parameters
- Hepatitis
- Hirschsprung's disease
- HIV-1 control
- Homocysteine levels
- Idiopathic pulmonary fibrosis
- IgE levels
- Inflammatory bowel disease
- Intracranial aneurysm
- Iris color
- Iron status markers
- Ischemic stroke
- Juvenile idiopathic arthritis
- Kidney stones
- LDL cholesterol
- Leprosy
- Leptin receptor levels
- Liver enzymes
- LP (a) levels
- Lung cancer
- Major mood disorders
- Malaria
- Male pattern baldness
- Matrix metalloproteinase levels
- MCP-1
- Melanoma
- Menarche & menopause
- Multiple sclerosis
- Myeloproliferative neoplasms
- Narcolepsy
- Nasopharyngeal cancer
- Neuroblastoma
- Nicotine dependence
- Obesity
- Open personality
- Osteoarthritis
- Osteoporosis
- Otosclerosis
- Other metabolic traits
- Ovarian cancer
- Pain
- Pancreatic cancer
- Panic disorder
- Parkinson's disease
- Periodontitis
- Peripheral arterial disease
- Phosphatidylcholine levels
- Platelet count
- Primary biliary cirrhosis
- PR interval
- Prostate cancer
- Protein levels
- Psoriasis
- Pulmonary funct. COPD
- QRS interval
- QT interval
- Quantitative traits
- Recombination rate
- Red vs.non-red hair
- Renal function
- Response to antipsychotic therapy
- Response to hepatitis C treatment
- Response to statin therapy
- Restless legs syndrome
- Rheumatoid arthritis
- Schizophrenia
- Serum metabolites
- Skin pigmentation
- Speech perception
- Sphingolipid levels
- Statin-induced myopathy
- Stroke
- Systemic lupus erythematosus
- Telomere length
- Testicular germ cell tumor
- Thyroid cancer
- Tooth development
- Total cholesterol
- Triglycerides
- Type 1 diabetes
- Type 2 diabetes
- Ulcerative colitis
- Urate
- Venous thromboembolism
- Vitamin B12 levels
- Warfarin dose
- Weight
- White cell count
- YKL-40 levels

Table 1 | Estimates of heritability and number of loci for several complex traits

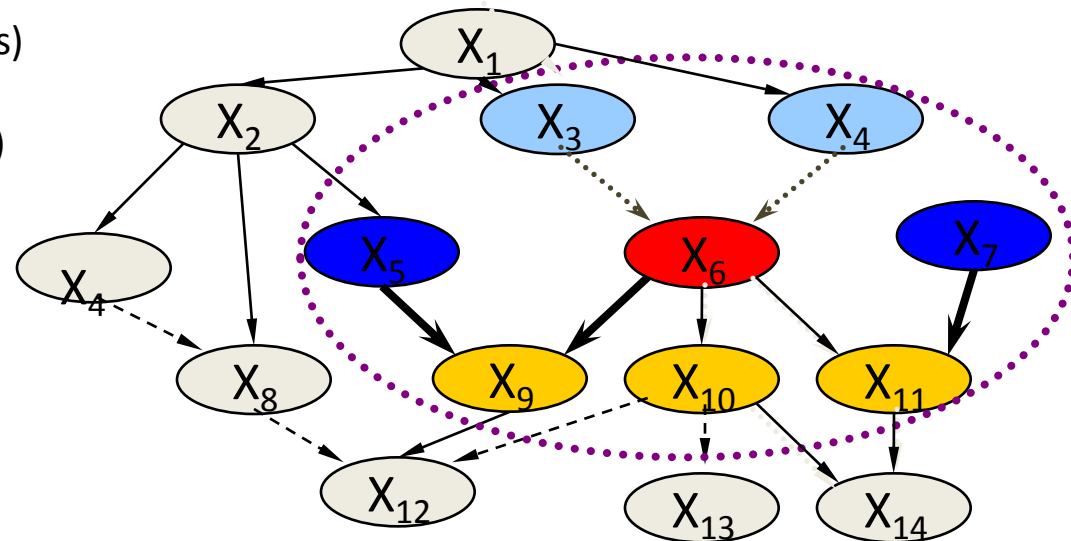
Disease	Number of loci	Proportion of heritability explained	Heritability measure
Age-related macular degeneration ⁷²	5	50%	Sibling recurrence risk
Crohn's disease ²¹	32	20%	Genetic risk (liability)
Systemic lupus erythematosus ⁷³	6	15%	Sibling recurrence risk
Type 2 diabetes ⁷⁴	18	6%	Sibling recurrence risk
HDL cholesterol ⁷⁵	7	5.2%	Residual* phenotypic variance
Height ¹⁵	40	5%	Phenotypic variance
Early onset myocardial infarction ⁷⁶	9	2.8%	Phenotypic variance
Fasting glucose ⁷⁷	4	1.5%	Phenotypic variance

* Residual is after adjustment for age, gender, diabetes.

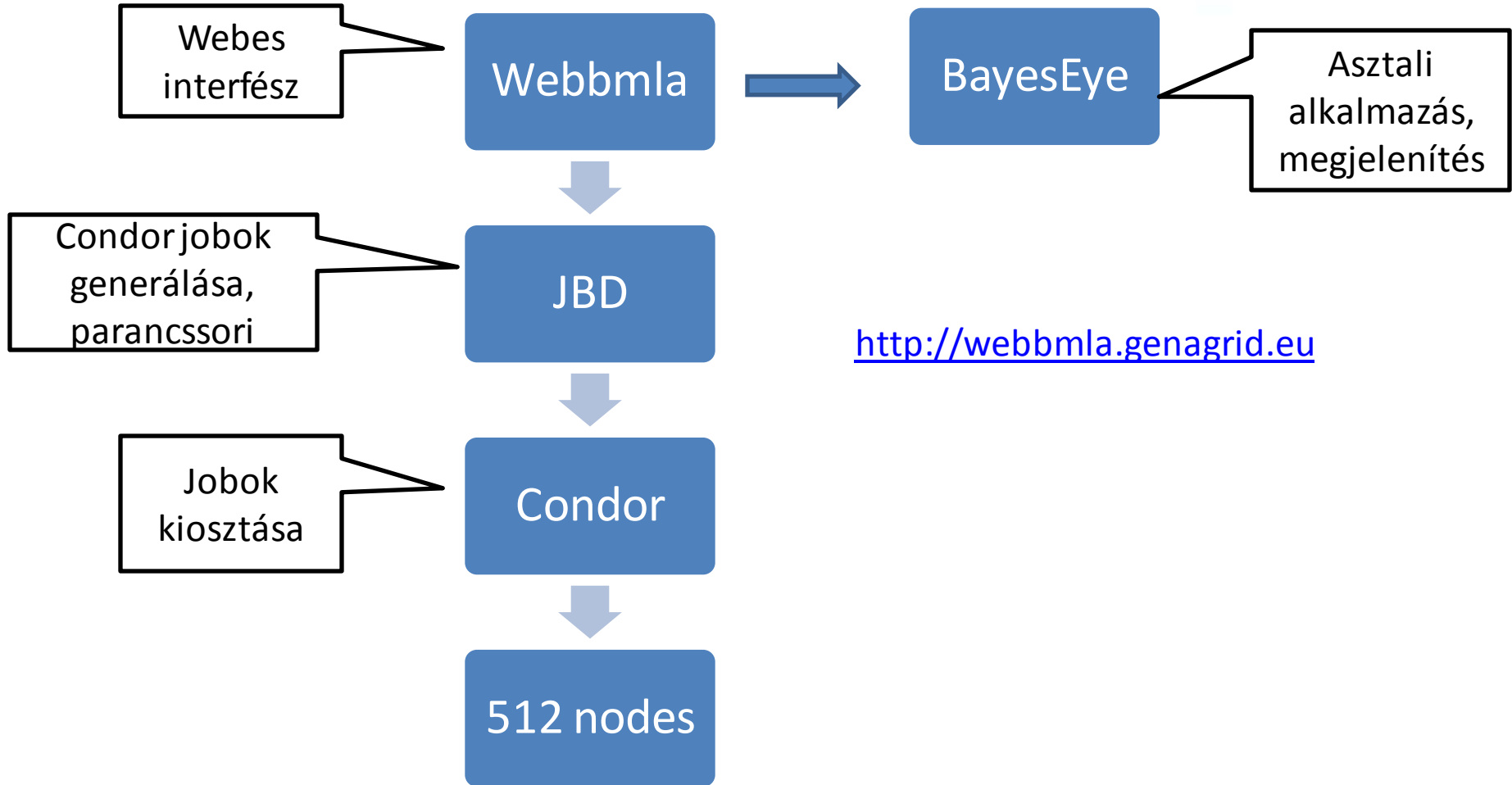
Részletesebb „nyelv” az asszociáció kutatásban gráf modellek felhasználásával



- Markov-takaró
- Számított jegyek:
 - Markov Blanket Membership (bináris)
 - Markov Blanket Set (halmaz)
 - Markov Blanket subgraph (struktúra)
- Számítás:
 - Markov chain Monte-Carlo szimulációk
 - Kis tár- és memóriaigény
 - Számításintenzív
- Parameter study



Architektúra



Webbmla



User's home page

Jobs

Name	Date	Status
10	2011/02/17 14:17:12	ready

gozos

Logout

Menu

User's home

Bmla analysis

File conversion

NGS

Haplo-Bmla analysis



© copyright 2010 Genagrid project

design: styleshout.com

Webbmla



BMLA Analysis Options

DATASET

GWA_inputed_170dataset.csv successfully uploaded

Bmla analysis

File conversion

NGS

AGGREGATION

Use custom gene aggregation

PERMUTATION TEST

Allow permutation tests

Number of permutation tests:

TARGET VARIABLES

Number of target variables:

Select method to handle the target variables

Individually and keeping the others
 Individually and increasing the priority

OTHER OPTIONS

Allow sequential analysis

Select method to use in sequential analysis:

Select analysis levels:

ADVANCED OPTIONS

Whole machine
 Connect confidence

Select priority:



© copyright 2010 GenaGrid project
design: ityackout.com

PERMUTATION TEST

Allow permutation tests

Number of permutation tests:

Webbmla



BMLA Analysis Options

DATASET
GWA_inputed_100dataset.csv successfully uploaded

AGGREGATION
 Use custom gene aggregation

PERMUTATION TEST
 Allow permutation test
Number of permutation tests:

TARGET VARIABLES
Number of target variables:
Select method to handle the target variables:
 individually and keeping the others
 individually and excluding the others
 jointly

OTHER OPTIONS
 Allow sequential analysis
Select method to use in sequential analysis: Decomposition of data into groups with size 4
Select analysis levels: RDR, RES and RDG

ADVANCED OPTIONS
 Whole machine
 Convert confidence
Select priority: Medium

NKTH
© copyright 2010 GenaGrid project
design: ityackout.com

Webbmla



BMLA Analysis Options

DATASET
GWA_inputed_rtdataset.csv successfully uploaded

AGGREGATION
 Use custom gene aggregation

PERMUTATION TEST
 Allow permutation tests

TARGET V
Number of targets:
Select method to use in permutation analysis:
 Use others

OTHER OPTIONS
 Allow sequential analysis
Select method to use in sequential analysis: Decomposition of data into groups with size 4
Select analysis levels: RDR, RES and RDG

ADVANCED OPTIONS
 Whole machine
 Connect confidence
Select priority: Medium

NKTH
© copyright 2010 GenaGrid project
design: ityackout.com

Webbmla



Analysis successfully started

The application is currently running. The results will be sent to you via email.
If you have any questions or remarks feel free to send it to us.

gozos Logout

Menu

User's home

Bmla analysis

File conversion

NGS



© copyright 2010 Genagrid project

design: styleshout.com

Webbmla



User's home page

Jobs

Name	Date	Status
10	2011/02/17 14:17:12	ready
proba	2011/03/22 10:06:24	run

gozos

[Logout](#)

Menu

[User's home](#)

[Bmla analysis](#)

[File conversion](#)

[NGS](#)

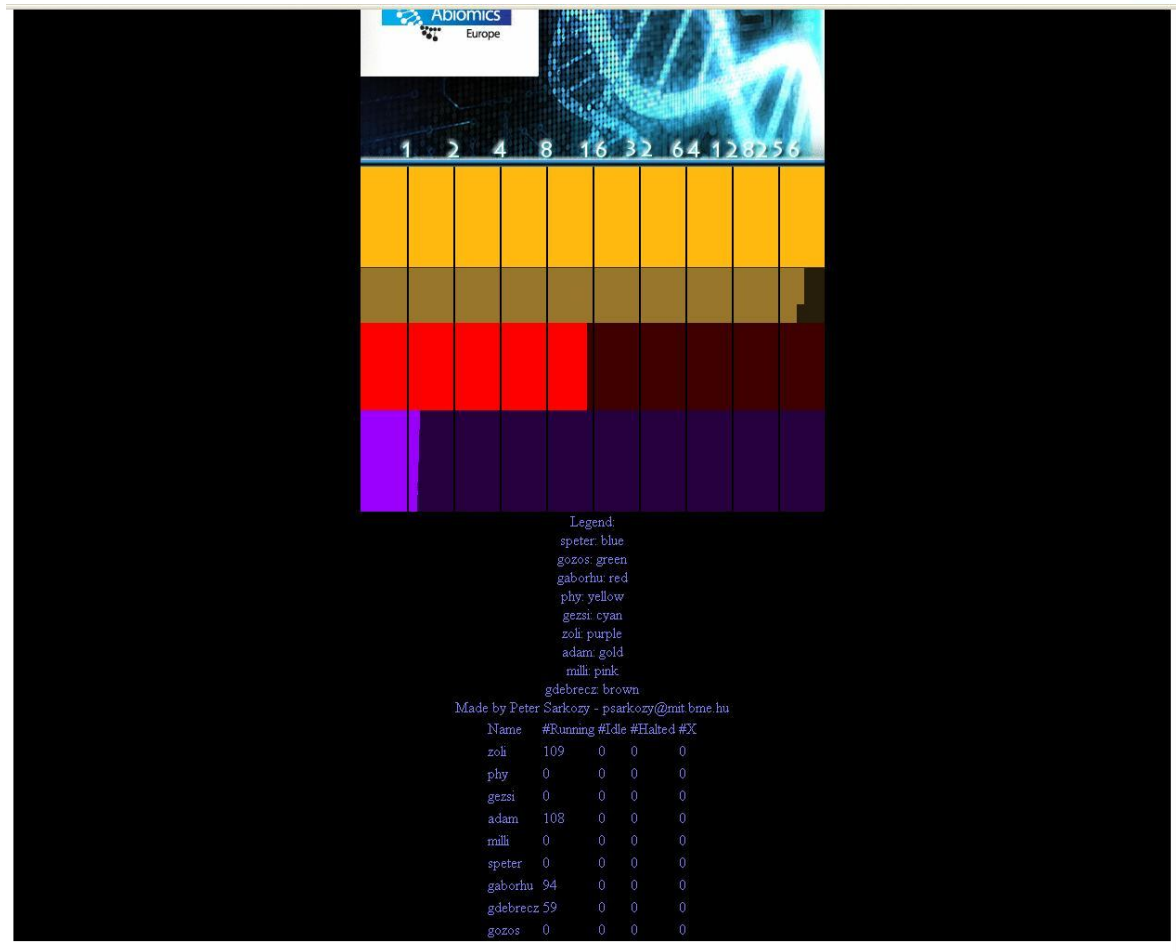
[Haplo-Bmla analysis](#)



© copyright 2010 **Genagrid project**

design: [styleshout.com](#)

Státusz



BayesEye



Content of the package

Use the filter combo boxes to find the results in the table that you desire and doubleclick it to open in a new editor window.

model	p.	cle...	ra...	p..	v..	v..	vssRati...	structPrior	data
/usr/people/gaborhu/GG_Asztma/Asztma_2010okt20/GG_Asztma_II_p1_model_2010...	7	false	true	CH	-1	-1	-1.000000	globalUniform	/usr/people/gaborhu/GG_Asztma/Asztma_2010c
/usr/people/gaborhu/GG_Asztma/Asztma_2010okt20/GG_Asztma_II_p1_model_2010...	7	false	true	CH	-1	-1	-1.000000	globalUniform	/usr/people/gaborhu/GG_Asztma/Asztma_2010c
/usr/people/gaborhu/GG_Asztma/Asztma_2010okt20/GG_Asztma_II_p1_model_2010...	7	false	true	CH	-1	-1	-1.000000	globalUniform	/usr/people/gaborhu/GG_Asztma/Asztma_2010c
/usr/people/gaborhu/GG_Asztma/Asztma_2010okt20/GG_Asztma_II_p1_model_2010...	7	false	true	CH	-1	-1	-1.000000	globalUniform	/usr/people/gaborhu/GG_Asztma/Asztma_2010c
/usr/people/gaborhu/GG_Asztma/Asztma_2010okt20/GG_Asztma_II_p1_model_2010...	7	false	true	CH	-1	-1	-1.000000	globalUniform	/usr/people/gaborhu/GG_Asztma/Asztma_2010c

model: /usr/people/gaborhu/GG_Asztma/Asztma_2010okt20/GG_Asztma_II_p1_model_2010okt13.xml

parentCount: 7 clearModel: false randomizeModel: true

paramPrior: CH vss_CH: -1 vss_BD: -1 vssRatio_BD: -1.000000 structPrior: globalUniform

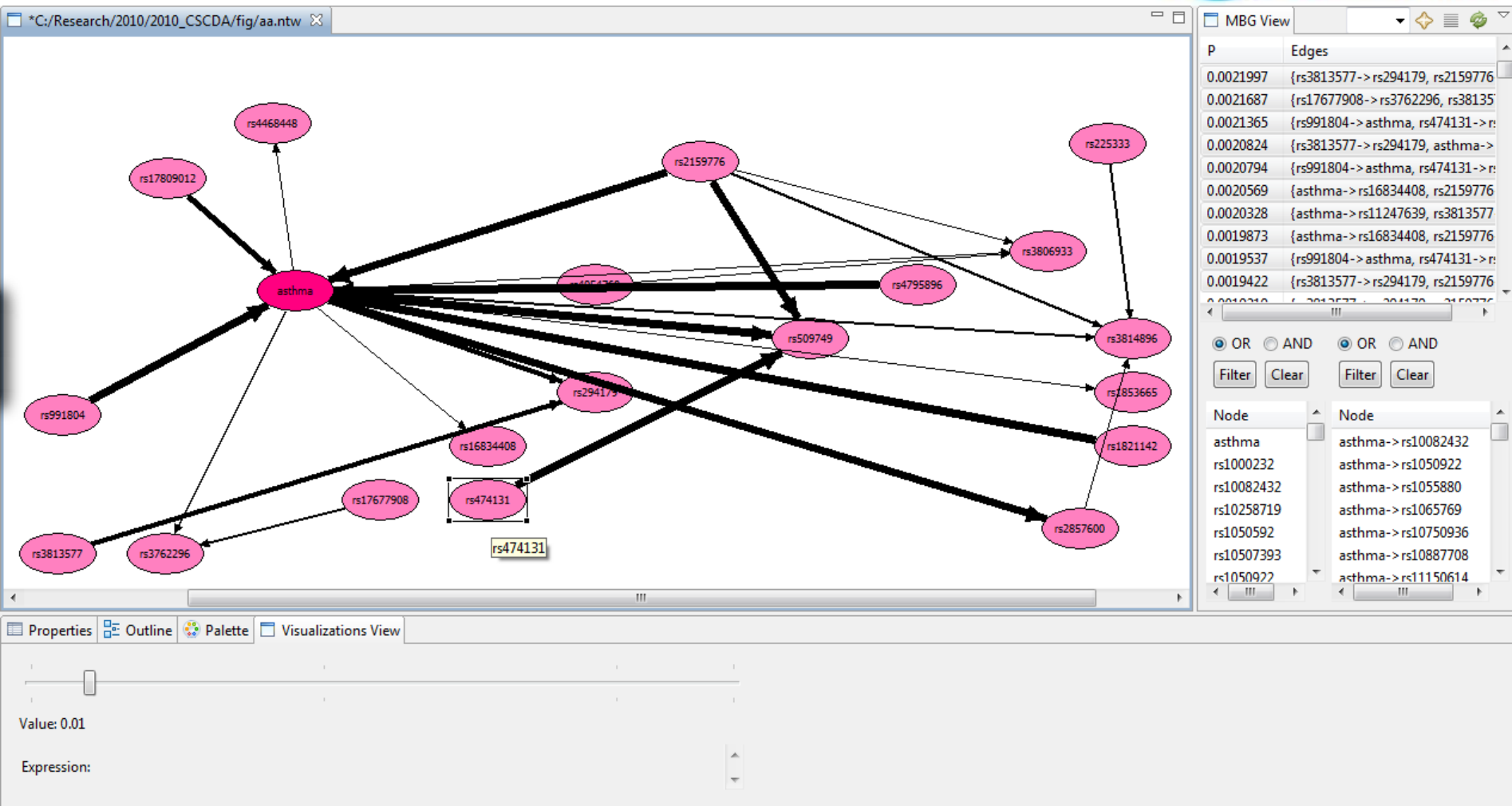
data: /usr/people/gaborhu/GG_Asztma/Asztma_2010okt20/GG_asztma_II_p1_data.csv sampleSize: 500 startIndex: 0 burnIn: 5000000 steps: 50000000

chainCount: 4 swapProfile: any heatProfile: hyperbolic temperature: 0.000000 mcmcMethod: dag

query: set_mbm_bid([asztma],X);set_mbs_bid([asztma],X);set_mbg_bid([asztma],X) randomSeed: convertIIC: true

Clear filters Close

BayesEye



sumProbMBG=0.12327

MBM Edges MBG MBS

Alkalmazások



- Petefészekrák, IOTA, Dirk Timmerman, 2003
- Asthma, Csaba Szalai, András Falus, 2007
- Allergia, Márta Széll, 2007
- Rheumatoid arthritis, Edit Buzás, 2008
- Leukémia, Csaba Szalai, András Falus, 2009
- Osteosarcoma, Csaba Szalai, András Falus, 2009
- Obezitás, pharma, 2009
- Heroin függés, Mária Sasvári, 2009
- Impulzivitás, Anna Székely, 2010
- Kognitív teljesítmény változása, Mária Sasvári, 2010
- Osteoporosis, Péter Lakatos, 2010
- Mucsi István, vesetranszplantáció
- Kiss István, dialízis
- Mátyus Péter, mellékhatás alapú gyógyszer újrapozicionálás (farmakogenomika)

- Sum: egy számítás 20 – 200 job, jobonként 12 – 200 óra terhelés

Munkatársak



- Sárközy Péter: genomszekvenálás, genotipizálás, haplotípuselemzés
- Gézsi András (SOTE-PhD): teljes genom szélességű vizsgálatok
- Zoltán Balázs (MSc): grid futtatás, GUI
- Hajós Gergely: szekvenciális kísérlettervezés, grid
- Hullám Gábor: oksági indukció
- Millinghoffer András: MCMC módszerek és eredmények interpretációja
- Temesi Gergely (SOTE-PhD): Adat- és tudásintegráció
- Marx Péter: folytonos célváltozókhoz jegykiválasztás
- Arany Ádám: farmakogenomika, szövegbányászat, grid, Bayes logika

Köszönetnyilvánítás



- OTKA PD-76348, MTA Bolyai János Research Fellowship (Antal P.),
- NKTH TECH-08-A1/2-2008-0120 (Genagrid),
- TÁMOP-4.2.1/B-09/1/KMR-2010-0002