

EDA in Pharmacogenomics

With
SAS JMP and Windows WF

Introduction

Hapmap, JMP and Windows WF

Hapmap JMP Windows WF Application

Next Steps

Conclusions

Questions

Dosing Models

Model Comparison

Modelling $\leftrightarrow$ EDA

**Statistical Workflow of a
Pharmacogenomic Study**

Warfarin

Subgroup Analysis

Warfarin clinical dosing algorithm		
	4.0376	
-	0.2546 x	Age in decades
+	0.0118 x	Height in cm
+	0.0134 x	Weight in kg
-	0.6752 x	Asian race
+	0.4060 x	Black or African American
+	0.0443 x	Missing or Mixed race
+	1.2799 x	Enzyme inducer status
-	0.5695 x	Amiodarone status
=	Square root of weekly warfarin dose**	

Demographic

Pharma

Clinical

From The International Warfarin Pharmacogenetics Consortium. N Engl J Med 2009;360(8):753-764

Pharmacogenetic Dosing Model

Warfarin pharmacogenetic dosing algorithm			
		5.6044	
-		0.2614 x	Age in decades
+		0.0087 x	Height in cm
+		0.0128 x	Weight in kg
-		0.8677 x	<i>VKORC1</i> A/G
-		1.6974 x	<i>VKORC1</i> A/A
-		0.4854 x	<i>VKORC1</i> genotype unknown
-		0.5211 x	<i>CYP2C9</i> *1/*2
-		0.9357 x	<i>CYP2C9</i> *1/*3
-		1.0616 x	<i>CYP2C9</i> *2/*2
-		1.9206 x	<i>CYP2C9</i> *2/*3
-		2.3312 x	<i>CYP2C9</i> *3/*3
-		0.2188 x	<i>CYP2C9</i> genotype unknown
-		0.1092 x	Asian race
-		0.2760 x	Black or African American
-		0.1032 x	Missing or Mixed race
+		1.1816 x	Enzyme inducer status
-		0.5503 x	Amiodarone status
=	Square root of weekly warfarin dose**		

Demographic

Genetic

Pharma

Clinical

From The International Warfarin Pharmacogenetics Consortium. N Engl J Med 2009;360(8):753-764

Table 2. Predicted Warfarin Doses with the Pharmacogenetic Algorithm, Clinical Algorithm, and Fixed-Dose Approach as Compared with the Actual Stable Dose in the Derivation and Validation Cohorts.*

Prediction Model	Derivation Cohort		Validation Cohort†	
	Mean Absolute Error (95% CI)		Mean Absolute Error (95% CI)	
	R ²		R ²	
	mg/wk	%	mg/wk	%
Pharmacogenetic algorithm‡§	8.3 (8.1–8.6)	47	8.5 (8.0–9.0)	43
Clinical algorithm§	10.0 (9.7–10.3)	27	9.9 (9.3–10.4)	26
Fixed-dose approach¶	13.3 (13.0–13.5)	0	13.0 (12.4–13.6)	0

* The 95% confidence intervals (CIs) on the estimates of mean absolute error were computed by bootstrapping with 1000 replications. R² is the coefficient of determination.

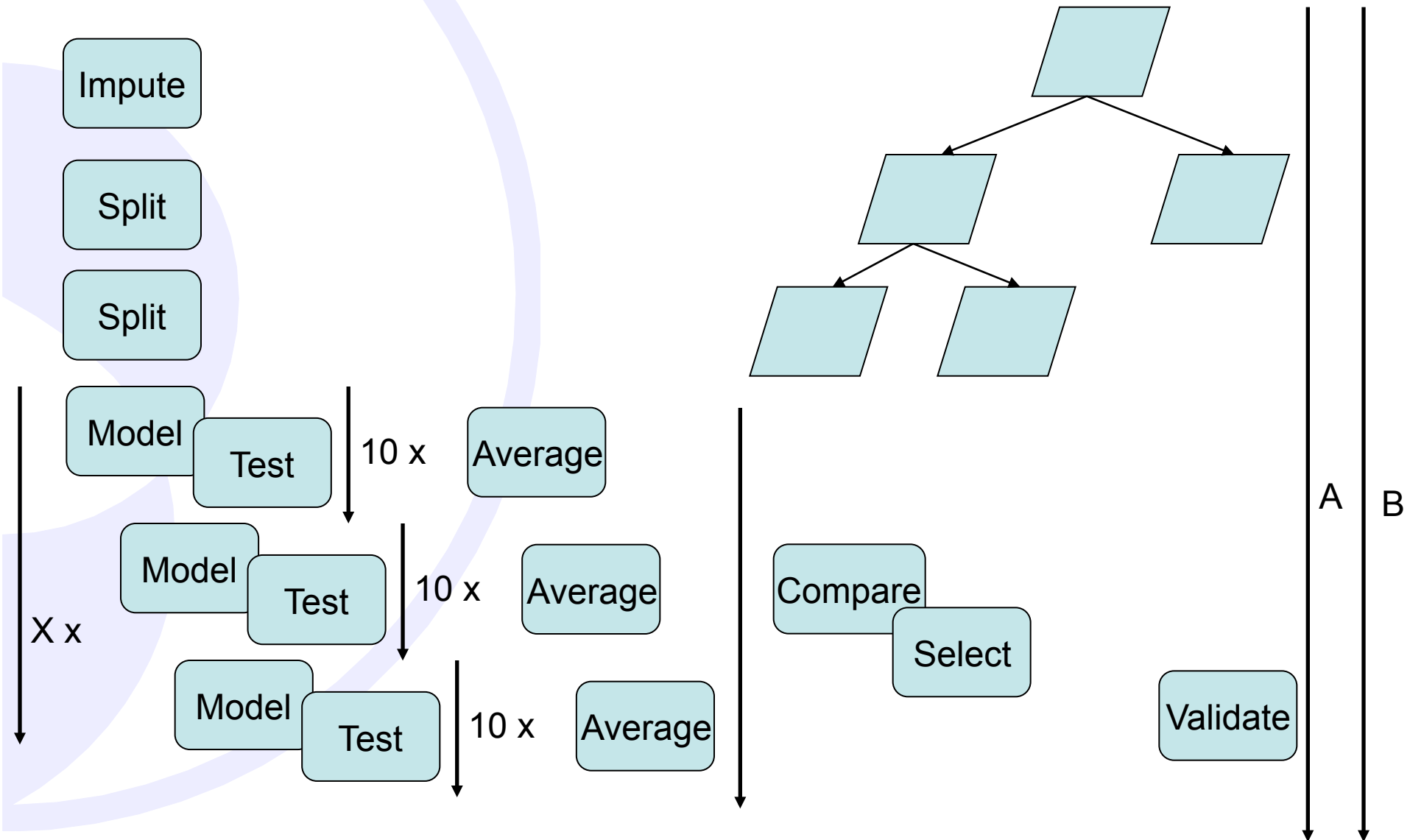
† In the calculation of the mean absolute error in the validation cohort, data from one patient who was taking an unusually high dose of warfarin were excluded. For details, see Section 11 in Supplementary Appendix 1.

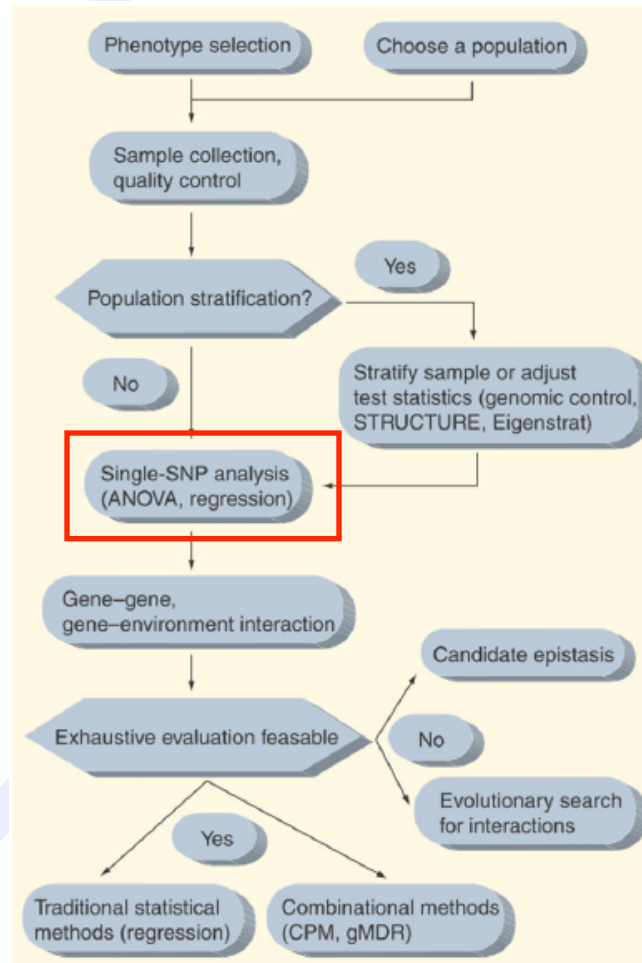
‡ P<0.001 for the pharmacogenetic algorithm as compared with the clinical algorithm, as derived with the use of McNemar's test of paired proportions.

§ P<0.001 for the pharmacogenetic algorithm as compared with the fixed-dose approach and for the clinical algorithm as compared with the fixed-dose approach, as derived with the use of McNemar's test of paired proportions.

¶ The fixed dose was 35 mg of warfarin per week.

From The International Warfarin Pharmacogenetics Consortium. N Engl J Med 2009;360(8):753-764





From Turner SD, Crawford DC, Ritchie MD. Expert Rev Clin Pharmacol 2009;2(5):559-570

Modelling / Comparisons: Routine Process

- WF compatible

EDA : Investigative Process

- Smaller (routine) processes

Model Extension

- Modelling + EDA
- WF possibility ?

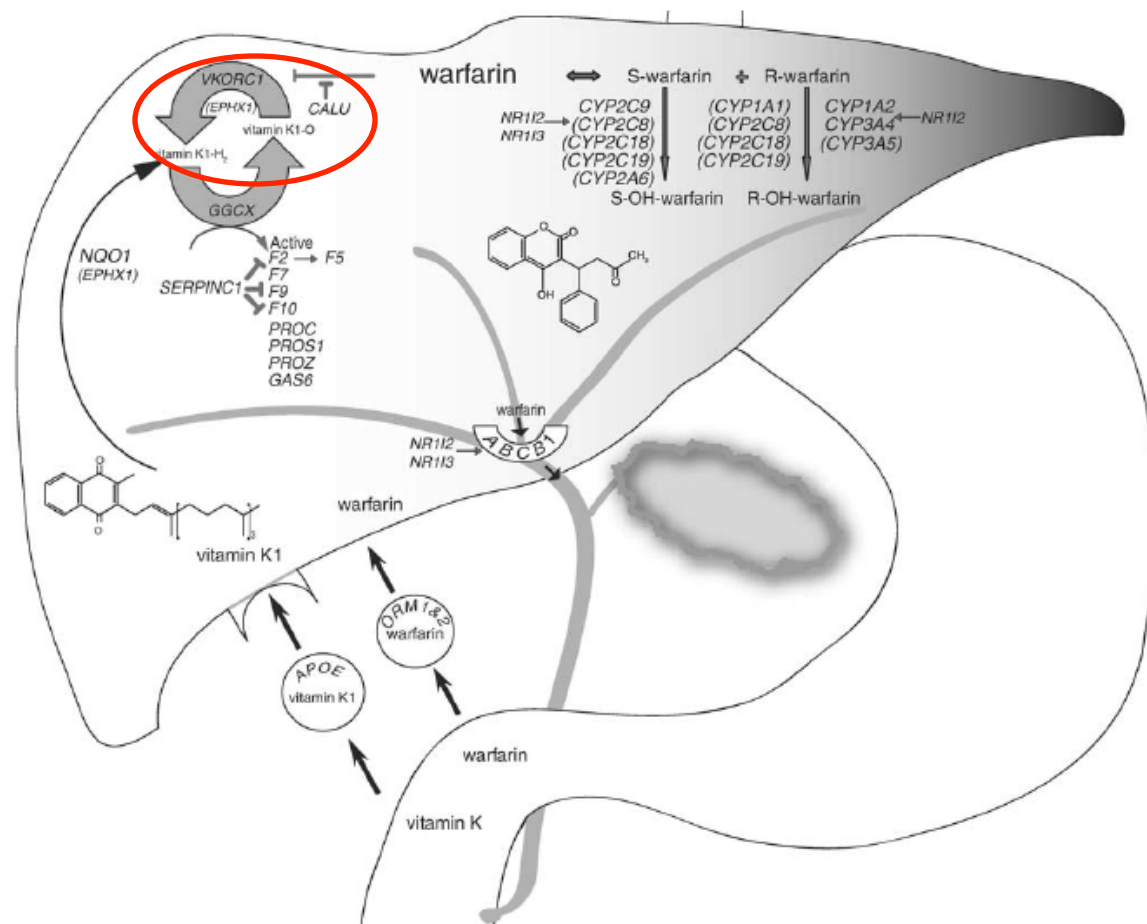


Figure 1 An overview of warfarin interactive pathways. This figure illustrates the genes thought to be involved in the action and biotransformation of warfarin and vitamin K.

From Wadelius M, Pirmohamed M. Pharmacogenomics J 2006;7(2): 99-111.

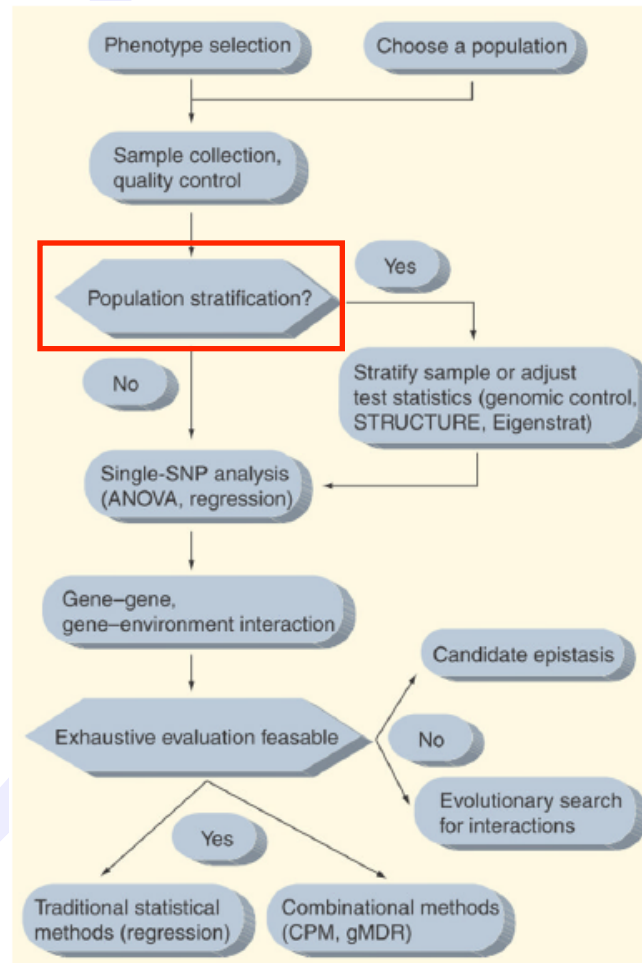
Table 1 Genes involved in the mechanism of action of warfarin

Protein name	Gene	Location	Exons	Transcript (bp)	Protein (aa)	Function of protein
Biotransformation of warfarin						
Transport						
Alpha-1-acid glycoprotein 1, Orosomucoid 1	ORM1	Chr 9: 114 083 890–114 087 309bp	6	802	201	A plasma glycoprotein that functions as a carrier of warfarin in the blood ^{16,17}
Alpha-1-acid glycoprotein 2, Orosomucoid 2	ORM2	Chr 9: 114 171 703–114 175 086bp	6	760	201	A plasma glycoprotein that functions as a carrier of warfarin in the blood ^{16,17}
P-glycoprotein, Multidrug resistance protein 1	ABCB1 (MDR1)	Chr 7: 85 668 428–85 877 818bp	29	4643	1279	A cellular efflux pump for xenobiotics. ²⁰ Warfarin is a weak inhibitor and maybe a substrate ¹⁸

Table 1 Continued

Protein name	Gene	Location	Exons	Transcript (bp)	Protein (aa)	Function of protein
Biotransformation of vitamin K						
Transport						
Apolipoprotein E	APOE	Chr 19: 50 100 879–50 104 489bp	4	1179	317	Apolipoprotein E serves as a ligand for receptors that mediate the uptake of vitamin K ^{61–64}
Vitamin K cycle						
Vitamin K epoxide reductase	VKORC1	Chr 16: 31 009 677–31 013 777bp	3	997	163	A hepatic epoxide hydrolase that catalyses the reduction of vitamin K. The target of warfarin ^{5,6,69,70}
Epoxide hydrolase 1, microsomal	EPHX1	Chr 1: 222 304 587–222 339 995bp	9	1605	455	A hepatic epoxide hydrolase in the endoplasmic reticulum that may be complexed with VKOR ^{86,87,89}
NAD(P)H dehydrogenase, quinone 1	NQO1	Chr 16: 68 300 807–68 317 893bp	6	2448	274	A detoxifying enzyme that has the potential to reduce the quinone form of vitamin K ^{62,90,91}
Calumenin	CALU	Chr 7: 127 973 368–128 005 478bp	7	3316	315	Binds to the vitamin K epoxide reductase complex and inhibits the effect of warfarin ^{98,99}
Gamma-glutamyl carboxylase	GGCX	Chr 2: 85 687 865–85 700 237bp	15	3155	758	Carboxylates vitamin K-dependent coagulation factors and proteins in the vitamin K cycle ^{92–94}

From Wadelius M, Pirmohamed M. Pharmacogenomics J 2006;7(2): 99-111.



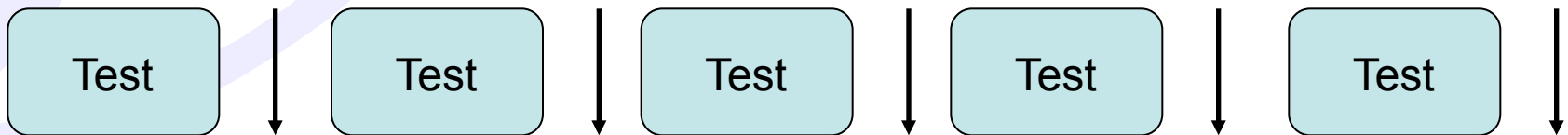
From Turner SD, Crawford DC, Ritchie MD. Expert Rev Clin Pharmacol 2009;2(5):559-570

Presumable more unknown factors are involved that are distributed differently in different populations, e.g. dietary habits.

Given this state of knowledge the question remains whether to use one universal dosing algorithm or to use different dosing algorithms for different subpopulations.

This subgroup analysis could be extended to other model factors

- clinical (diseases / disease states)
- pharma (high responder / low responder)



Hapmap

JMP

Hapmap JMP Workflow

Windows WF

Windows WF Activities

Activities inside a State



International HapMap Project

[Home](#) | [About the Project](#) | [Data](#) | [Publications](#) | [Tutorial](#)

[中文](#) | [English](#) | [Français](#) | [日本語](#) | [Yoruba](#)

The International HapMap Project is a partnership of scientists and funding agencies from Canada, China, Japan, Nigeria, the United Kingdom and the United States to develop a public resource that will help researchers find genes associated with human disease and response to pharmaceuticals. See "[About the International HapMap Project](#)" for more information.

Project Information

- [About the Project](#)
- [HapMap Publications](#)
- [HapMap Tutorial](#)
- [HapMap Mailing List](#)
- [HapMap Project Participants](#)
- [HapMap Mirror Site in Japan](#)

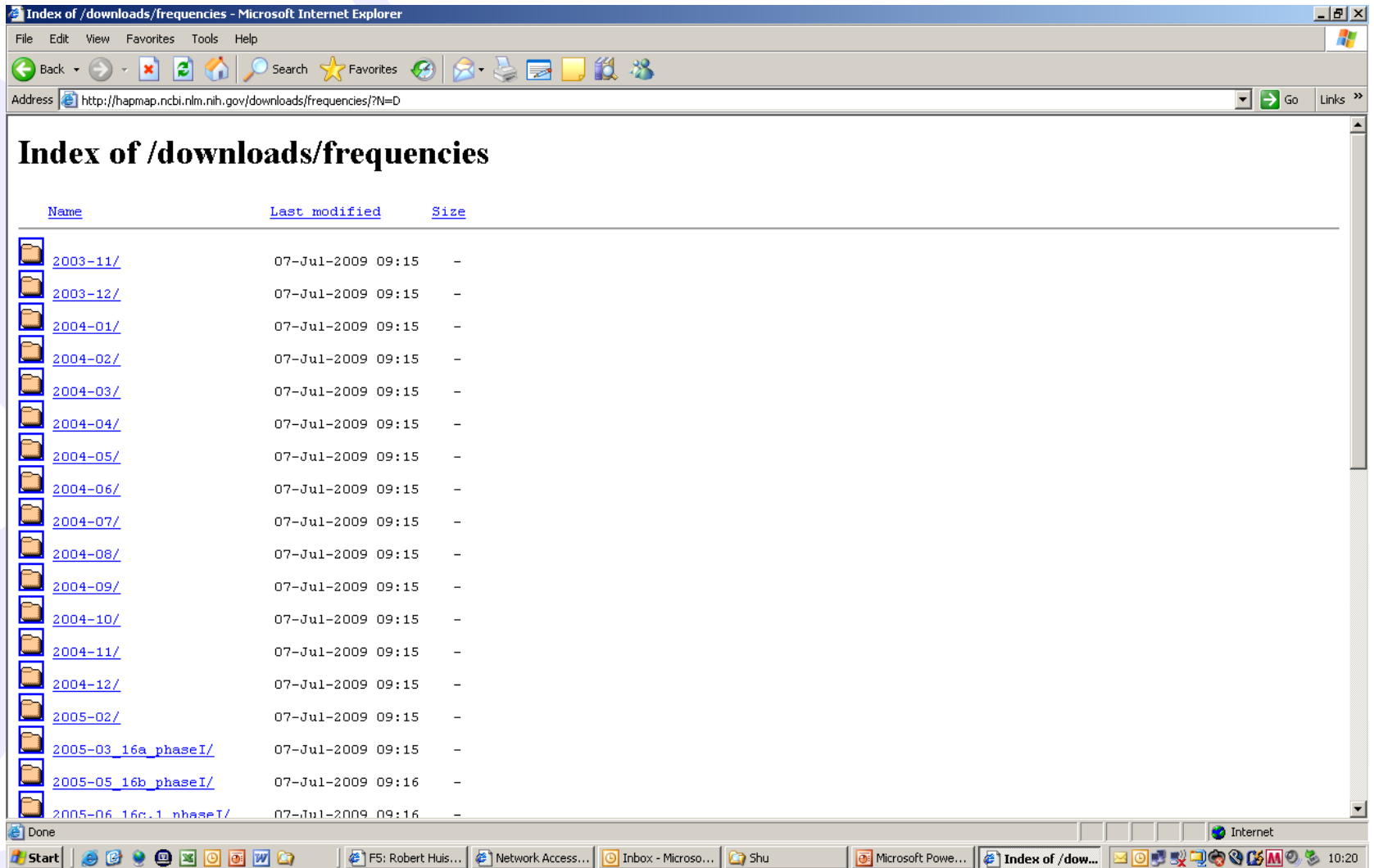
Project Data

- [HapMap Genome Browser \(Phase 1, 2 & 3 - merged genotypes & frequencies \)](#)
- [HapMap Genome Browser \(Phase 3 - genotypes, frequencies & LD \)](#)
- [HapMap Genome Browser \(Phase 1 & 2 - full dataset \)](#)
- [GWAs Karyogram](#)
- [HapMart](#)
- [HapMap FTP](#)

News

- 2009-12-14: Notice to Haploview users**
 Recently, there are several questions about Haploview data format errors, and these errors were observed when users tried to analyze HapMap release 27 data dumped from HapMap. The current Haploview version (4.1) does not work with release 27 data. Haploview will generate a software error similar to "Hapmap data format error: NA06984" when trying to open the data.

 The r27 data format will be supported by next Haploview version. There is a beta test version that is supposed to work and it can be obtained from <http://www.broadinstitute.org/haploview/haploview-downloads>. But since it is NOT an official release version, please use it base on your own judgment.
- 2009-12-10: Corrected HapMap3 phased haplotypes available for chromosome X**
 Phased haplotypes for consensus HapMap3 release 2 data for chromosome X has been corrected and the new data are now **available for bulk download**. Sorry for any inconvenience this might have caused.
- 2009-12-02: HapMap3 phased haplotypes available for chromosome X**
 Phased haplotypes for consensus HapMap3 release 2 data has been phased for chromosome X and are now available for bulk download. [Update: The downloading was disabled because several users have found that there are repeating data in some of the chrX phasing data files. The data source is being contacted and the downloading will be enabled as soon as the problem is



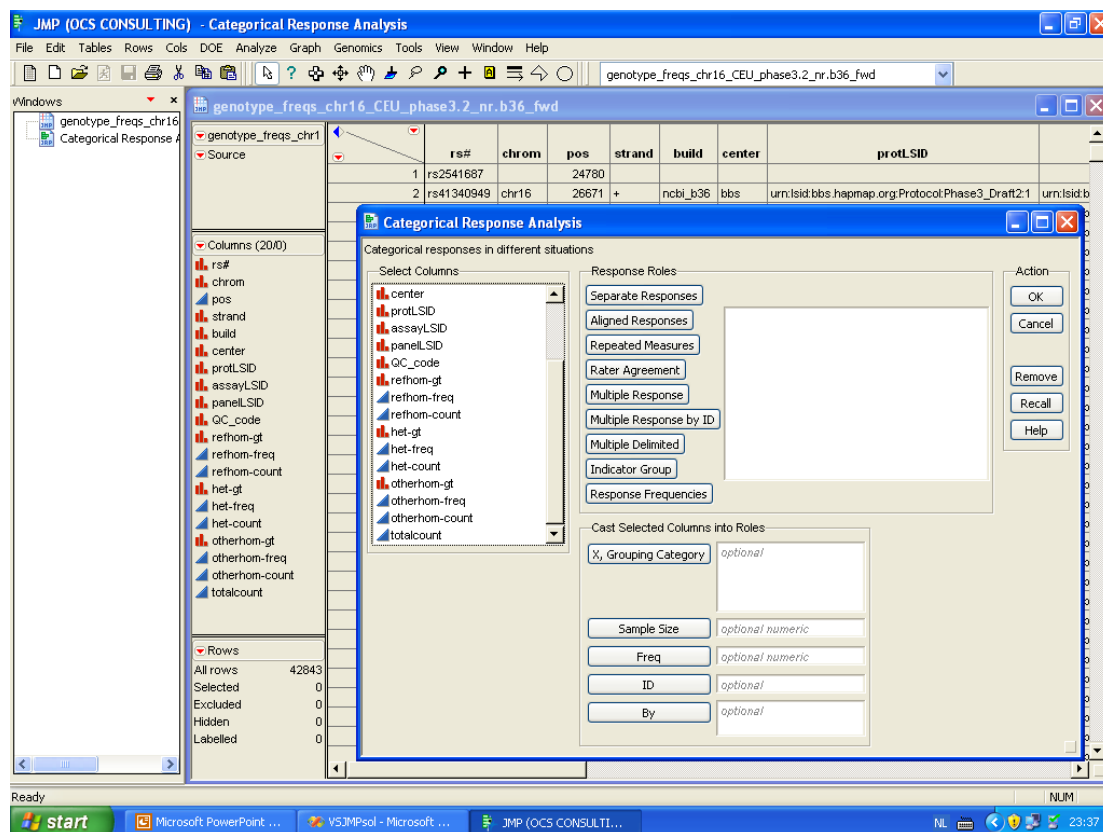
Index of /downloads/frequencies

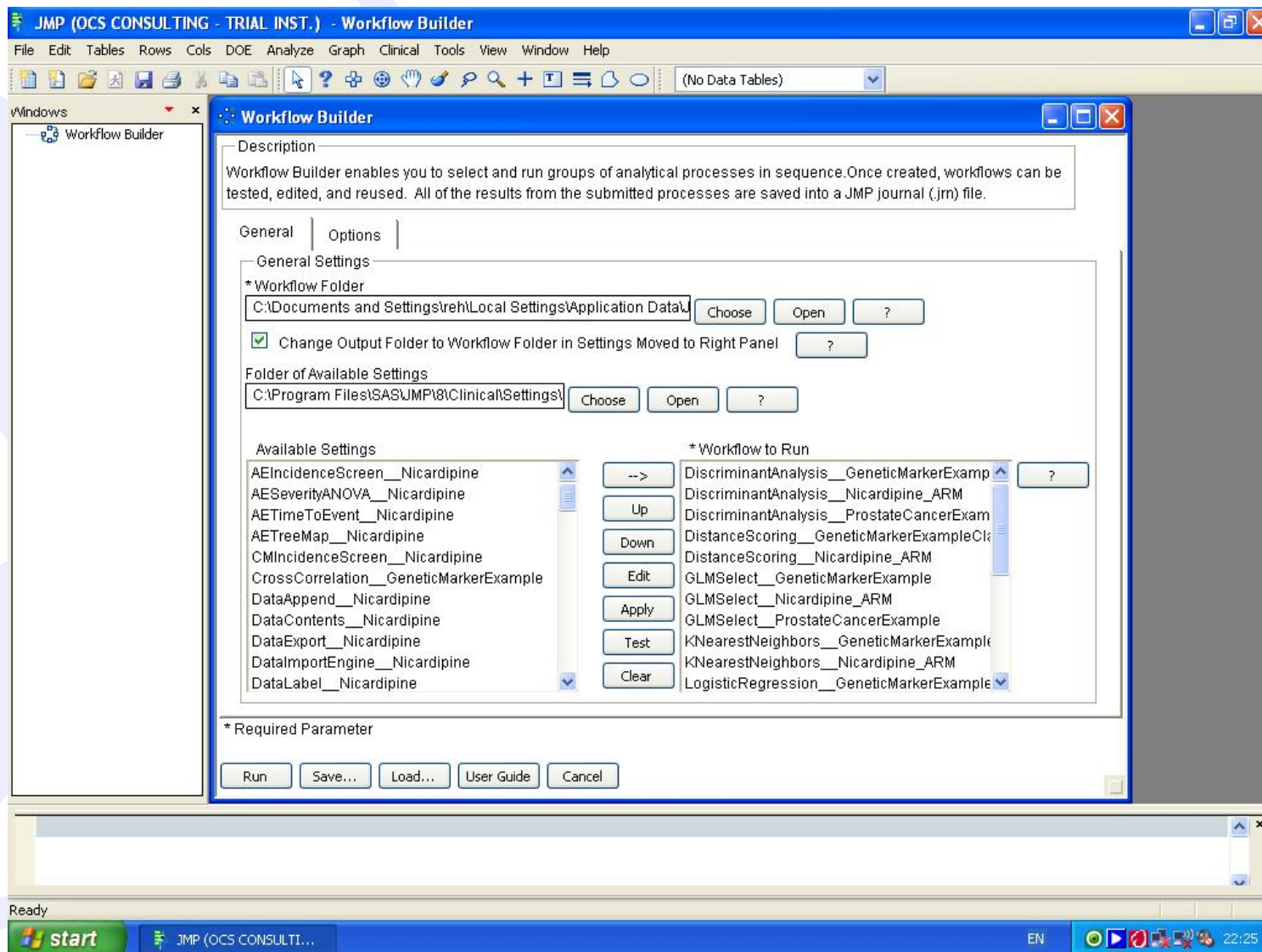
Name	Last modified	Size
2003-11/	07-Jul-2009 09:15	-
2003-12/	07-Jul-2009 09:15	-
2004-01/	07-Jul-2009 09:15	-
2004-02/	07-Jul-2009 09:15	-
2004-03/	07-Jul-2009 09:15	-
2004-04/	07-Jul-2009 09:15	-
2004-05/	07-Jul-2009 09:15	-
2004-06/	07-Jul-2009 09:15	-
2004-07/	07-Jul-2009 09:15	-
2004-08/	07-Jul-2009 09:15	-
2004-09/	07-Jul-2009 09:15	-
2004-10/	07-Jul-2009 09:15	-
2004-11/	07-Jul-2009 09:15	-
2004-12/	07-Jul-2009 09:15	-
2005-02/	07-Jul-2009 09:15	-
2005-03_16a_phaseI/	07-Jul-2009 09:15	-
2005-05_16b_phaseI/	07-Jul-2009 09:16	-
2005-06_16c.1_phaseI/	07-Jul-2009 09:16	-

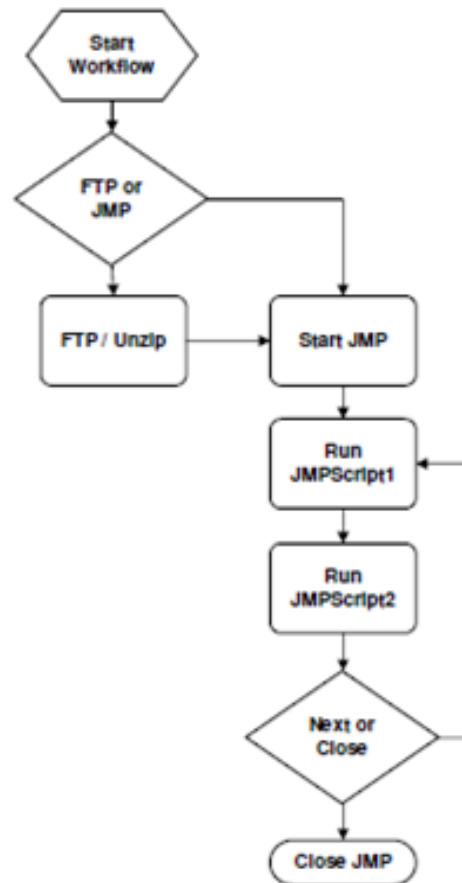
JMP Automation

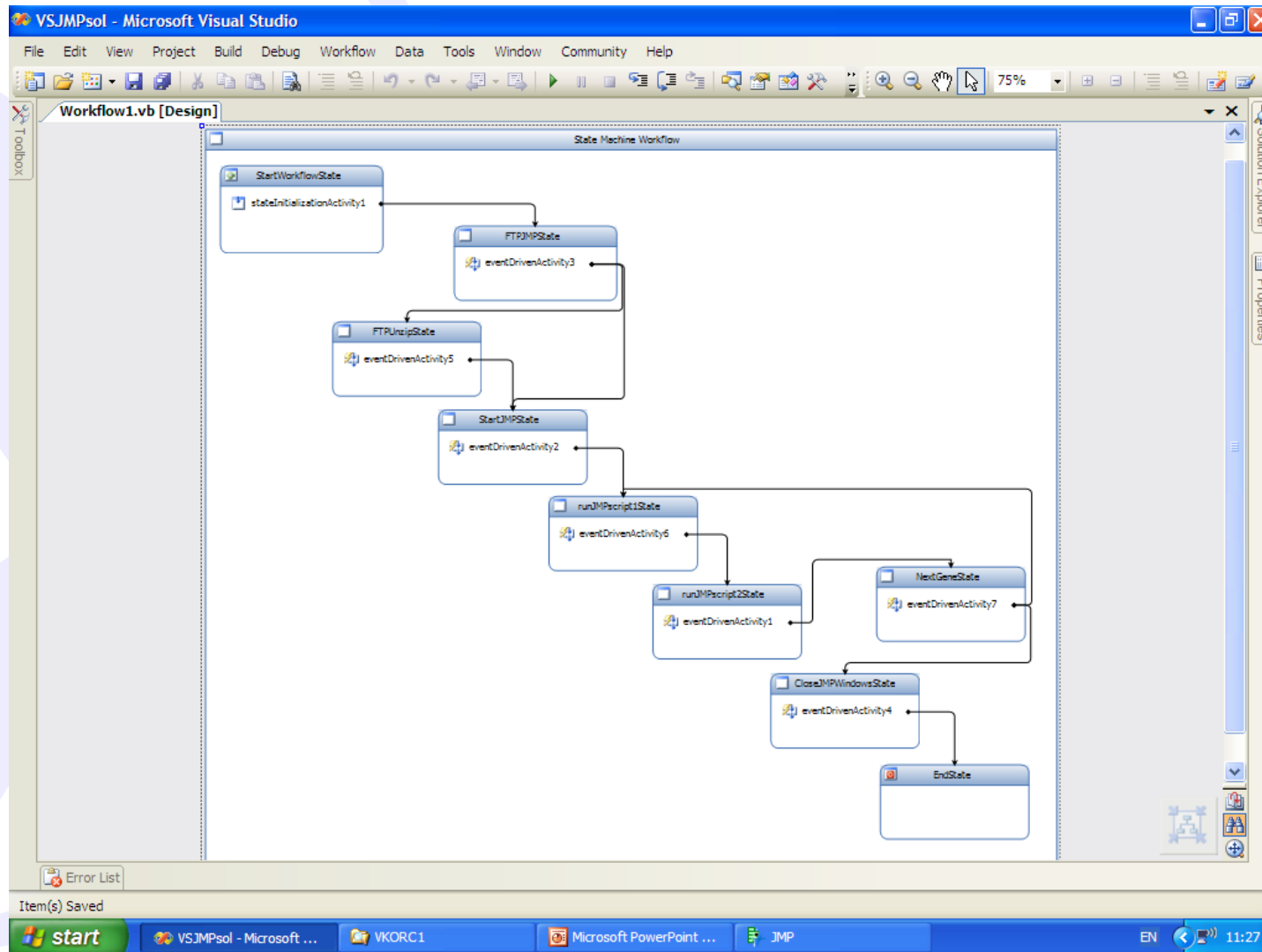
JMP Script

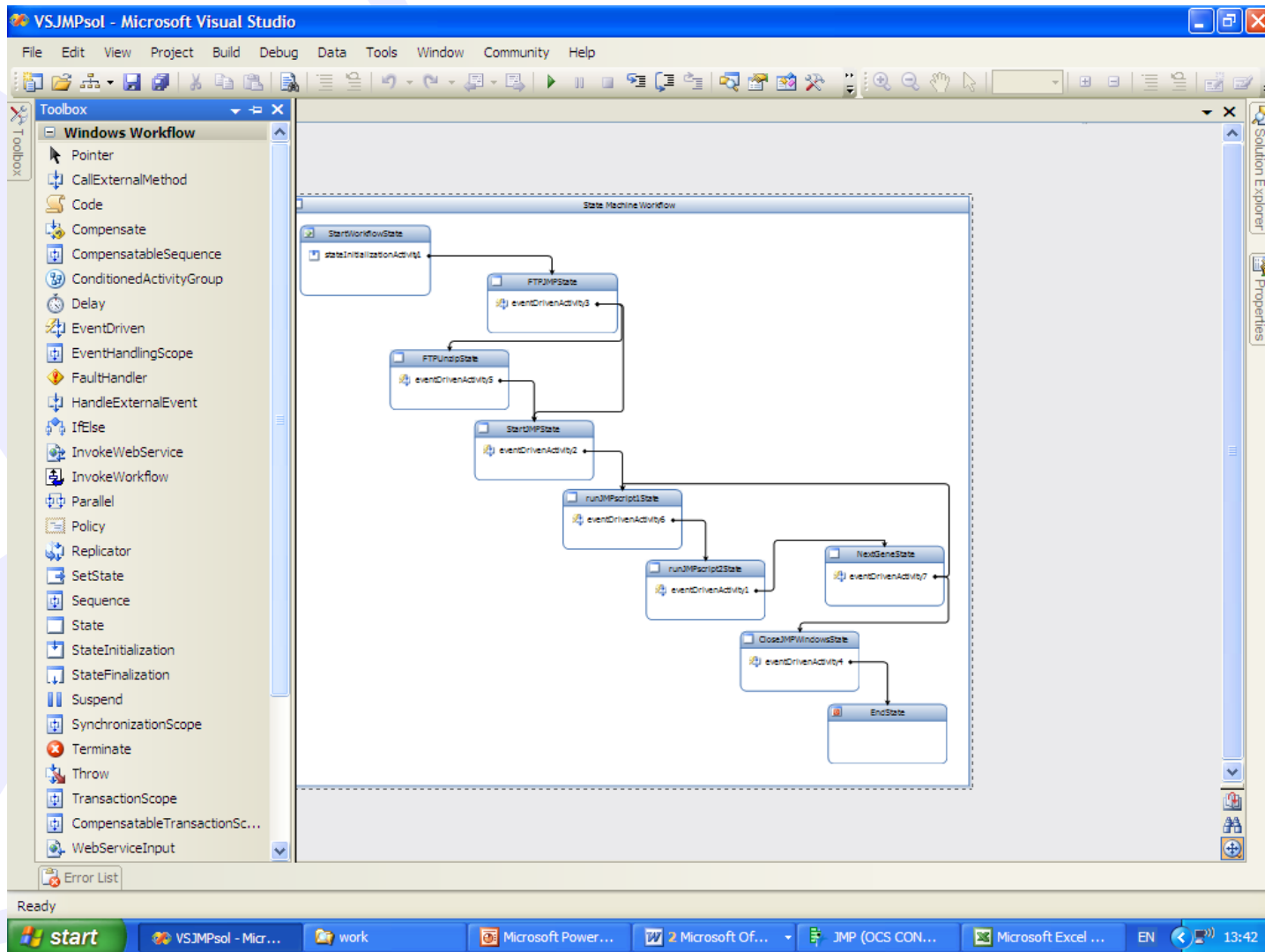
JMP Workflow Builder

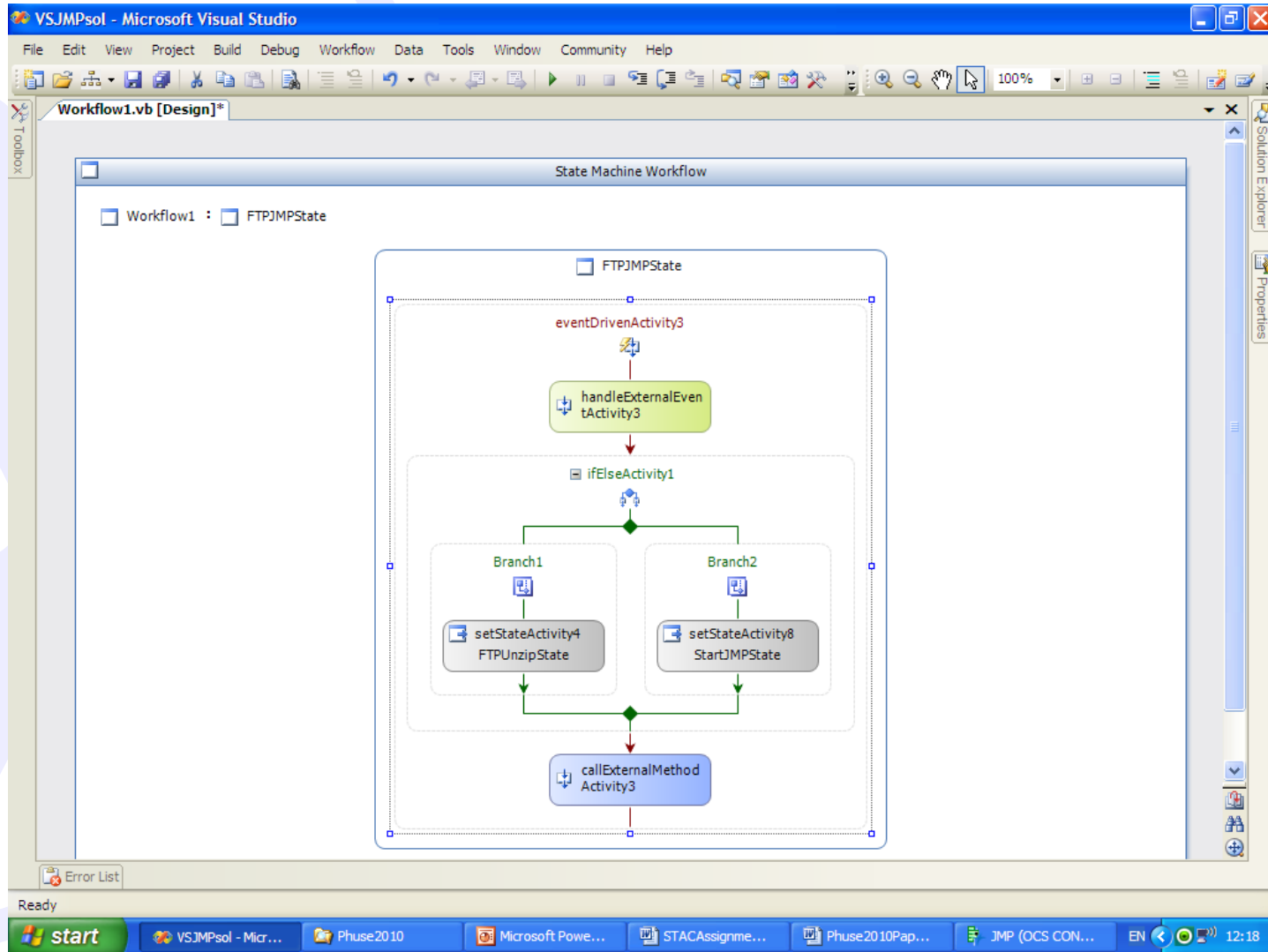


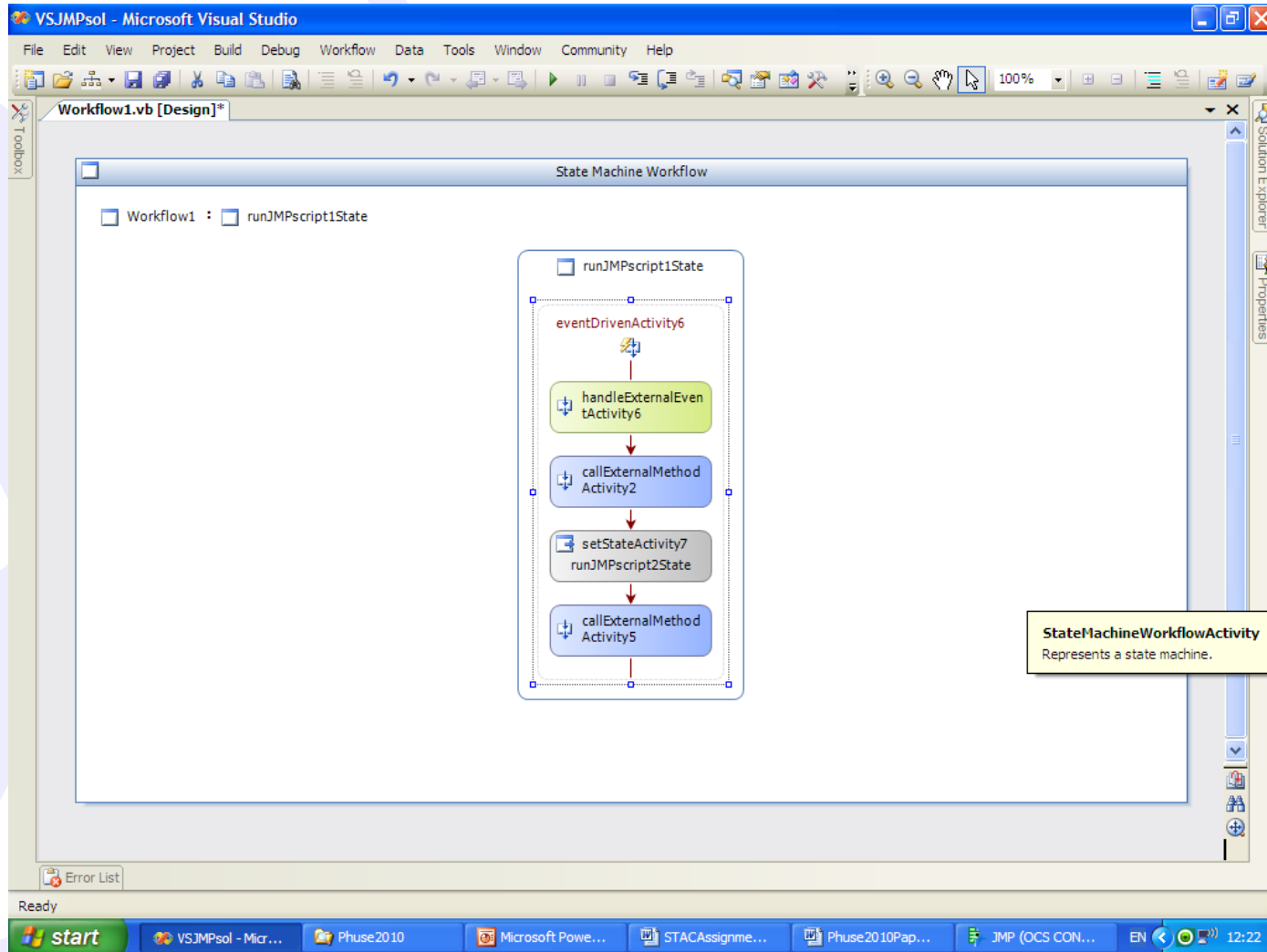












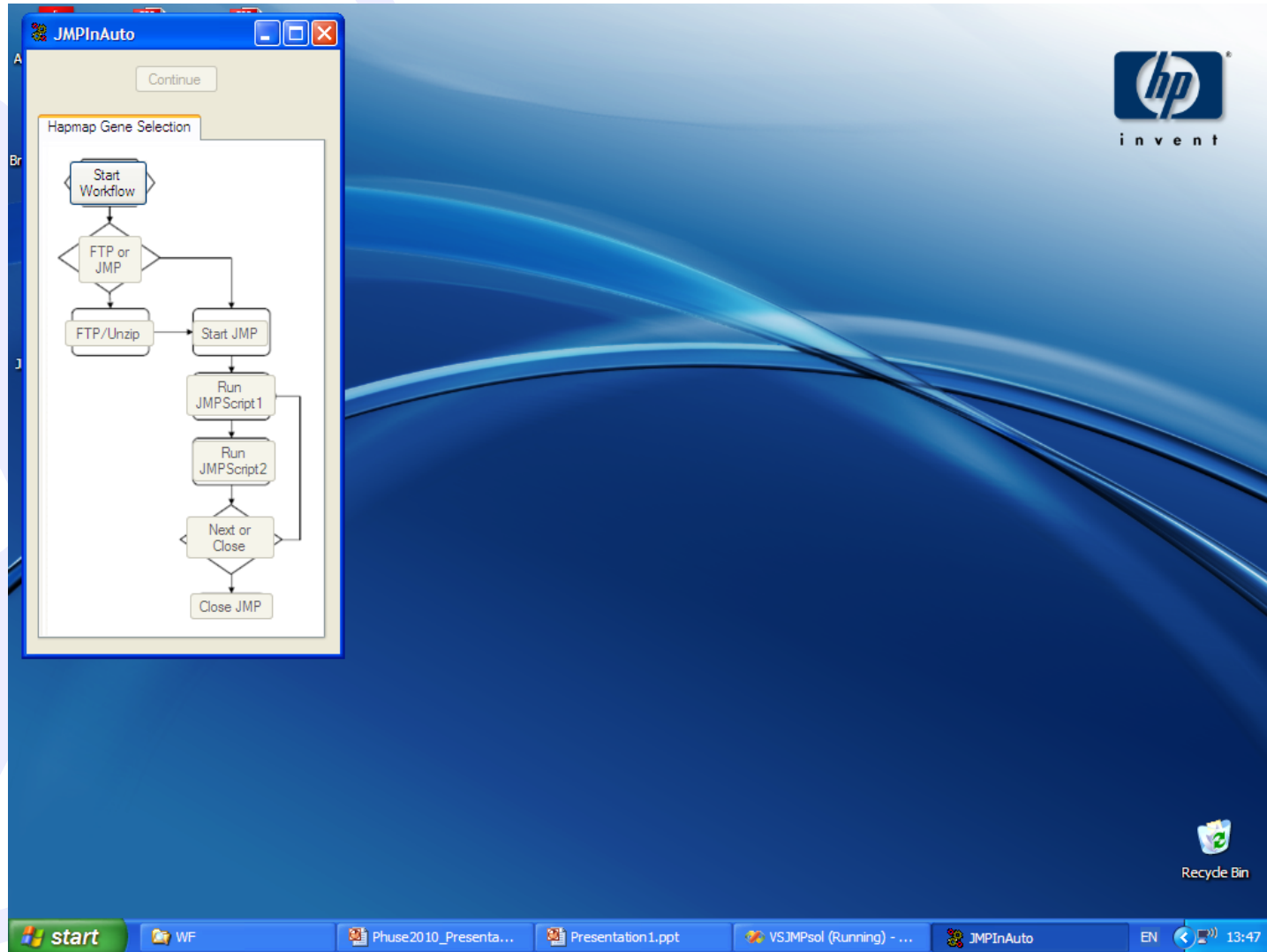
User Input for State Events

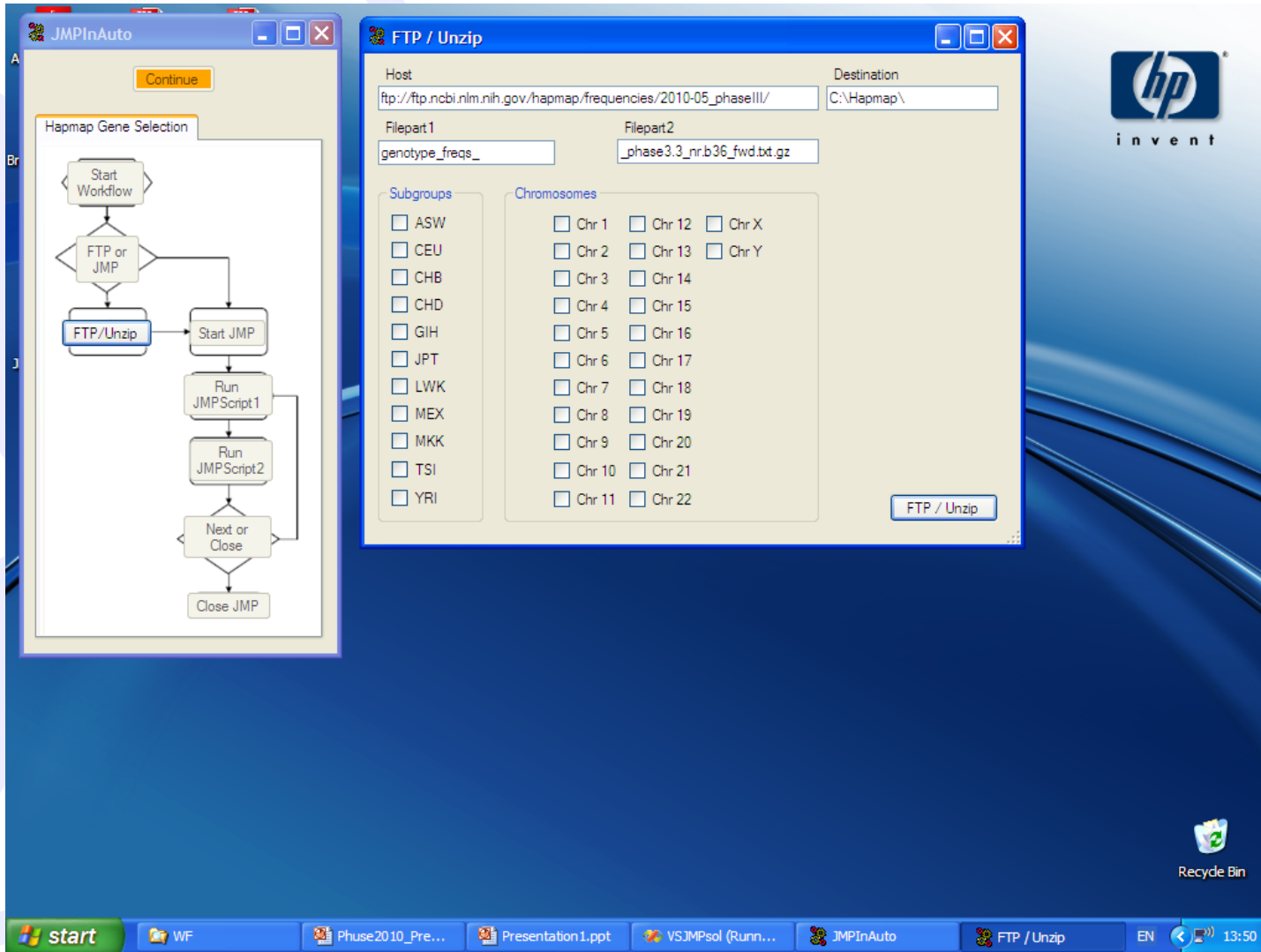
User Input for JMP Script

JMP Report

VKORC1 Marker rs9923231

JMP Script Results

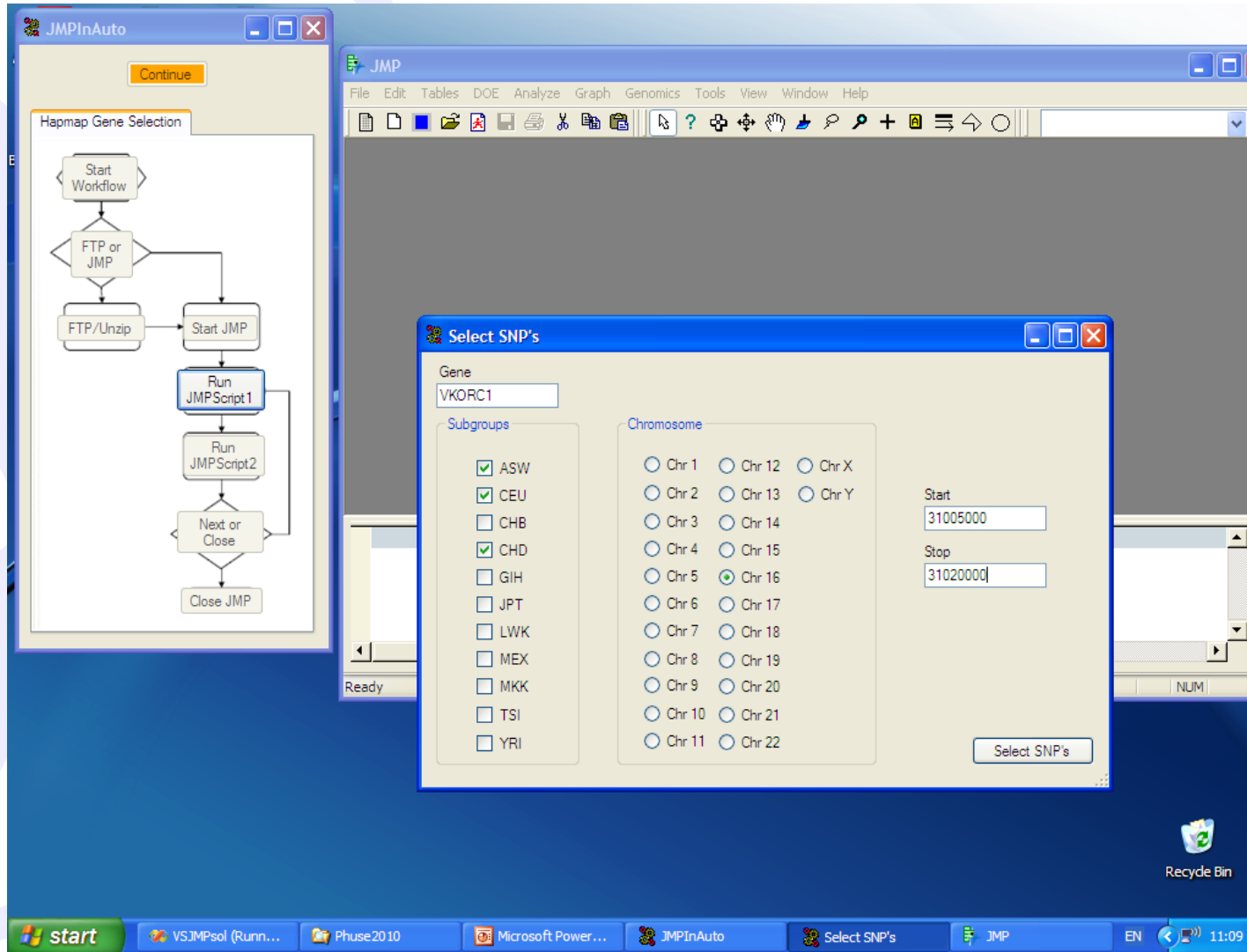




The screenshot displays a Windows desktop environment with two active windows. The background is a blue HP 'invent' logo wallpaper. The taskbar at the bottom shows the Start button and several open applications: 'WF', 'Phuse2010_Pre...', 'Presentation1.ppt', 'VSJMPsol (Runn...', 'JMPInAuto', and 'FTP / Unzip'. The system clock indicates 13:50 on EN.

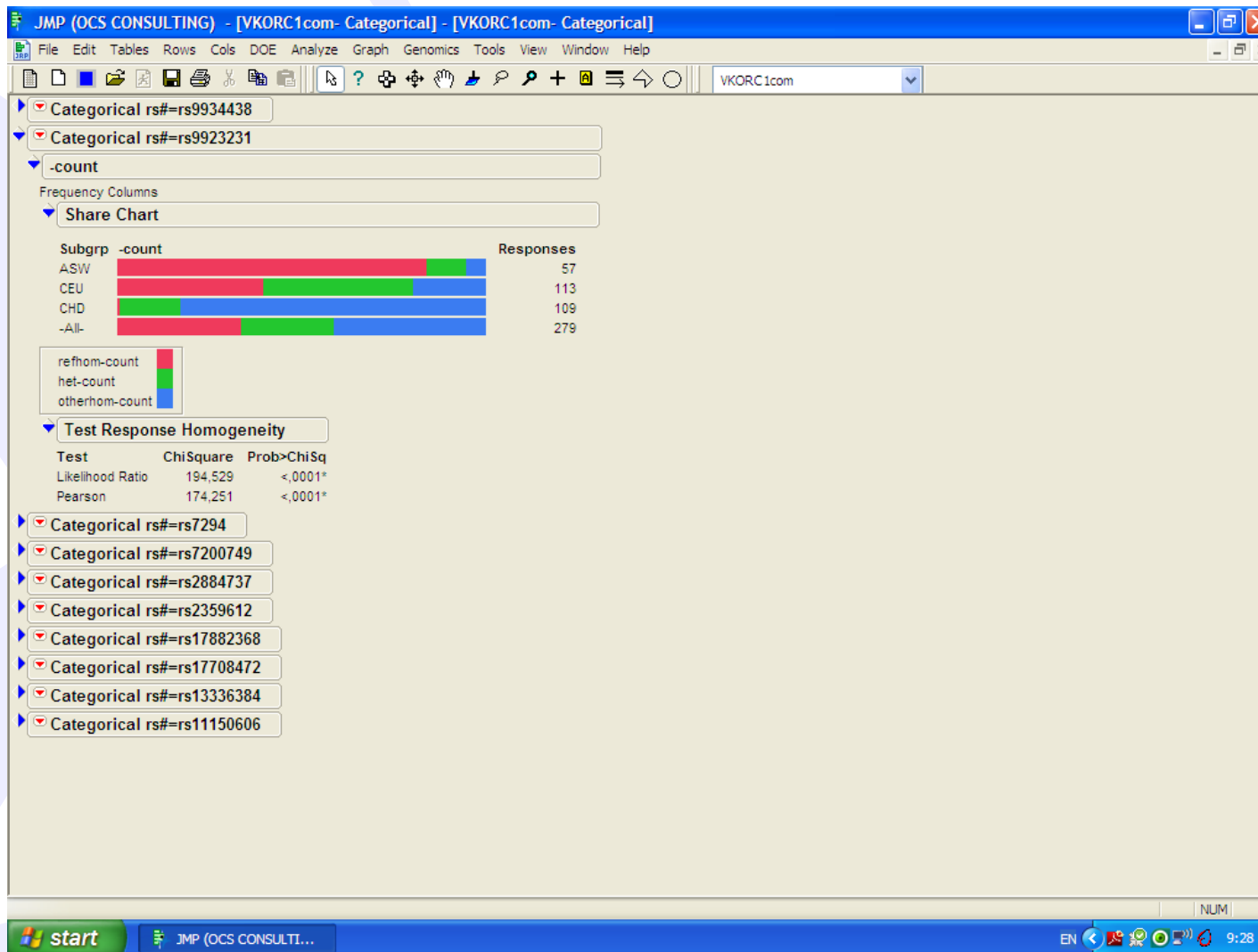
The **JMPInAuto** window features a 'Continue' button and a 'Hapmap Gene Selection' section. It contains a workflow diagram with the following steps: 'Start Workflow' (diamond), 'FTP or JMP' (diamond), 'FTP/Unzip' (rounded rectangle), 'Start JMP' (rounded rectangle), 'Run JMPScript1' (rounded rectangle), 'Run JMPScript2' (rounded rectangle), 'Next or Close' (diamond), and 'Close JMP' (rounded rectangle). Arrows indicate the flow from 'Start Workflow' through the 'FTP or JMP' decision to either 'FTP/Unzip' or 'Start JMP', then through the scripts to the 'Next or Close' decision, which leads to 'Close JMP'.

The **FTP / Unzip** window is a configuration dialog. It includes fields for 'Host' (ftp://ftp.ncbi.nlm.nih.gov/hapmap/frequencies/2010-05_phaseIII/) and 'Destination' (C:\Hapmap\). Below these are 'Filepart1' (genotype_freqs_) and 'Filepart2' (_phase3.3_nr.b36_fwd.bt.gz). There are two sections of checkboxes: 'Subgroups' (listing various populations like ASW, CEU, CHB, etc.) and 'Chromosomes' (listing chromosomes from Chr 1 to Chr 22, plus Chr X and Chr Y). An 'FTP / Unzip' button is located at the bottom right of the dialog.



The screenshot displays a Windows desktop environment with the following components:

- JMPInAuto Workflow:** A workflow window titled "Hapmap Gene Selection" showing a sequence of steps: "Start Workflow" → "FTP or JMP" → "FTP/Unzip" → "Start JMP" → "Run JMPScript1" → "Run JMPScript2" → "Next or Close" → "Close JMP". A "Continue" button is visible at the top.
- JMP Application:** The JMP software interface is open in the background, showing the standard menu bar (File, Edit, Tables, DOE, Analyze, Graph, Genomics, Tools, View, Window, Help) and a toolbar.
- Select SNP's Dialog:** A foreground dialog box for selecting SNPs. It includes:
 - Gene:** A text field containing "VKORC1".
 - Subgroups:** A list of population groups with checkboxes: ASW (checked), CEU (checked), CHB, CHD (checked), GIH, JPT, LWK, MEX, MKK, TSI, and YRI.
 - Chromosome:** A grid of radio buttons for chromosomes 1 through 22, plus Chr X and Chr Y. "Chr 16" is selected.
 - Start/Stop Coordinates:** Text fields for "Start" (31005000) and "Stop" (31020000).
 - Select SNP's Button:** A button at the bottom right to execute the selection.
- Taskbar:** The Windows taskbar at the bottom shows the Start button and several open applications: VSJMPsol (Runn...), Phuse2010, Microsoft Power..., JMPInAuto, Select SNP's, and JMP. The system clock shows 11:09.



VKORC1_rs9923231



rethom-count
 het-count
 otherhom-count

Test Response Homogeneity

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	194,529	<,0001*
Pearson	174,251	<,0001*



OCS

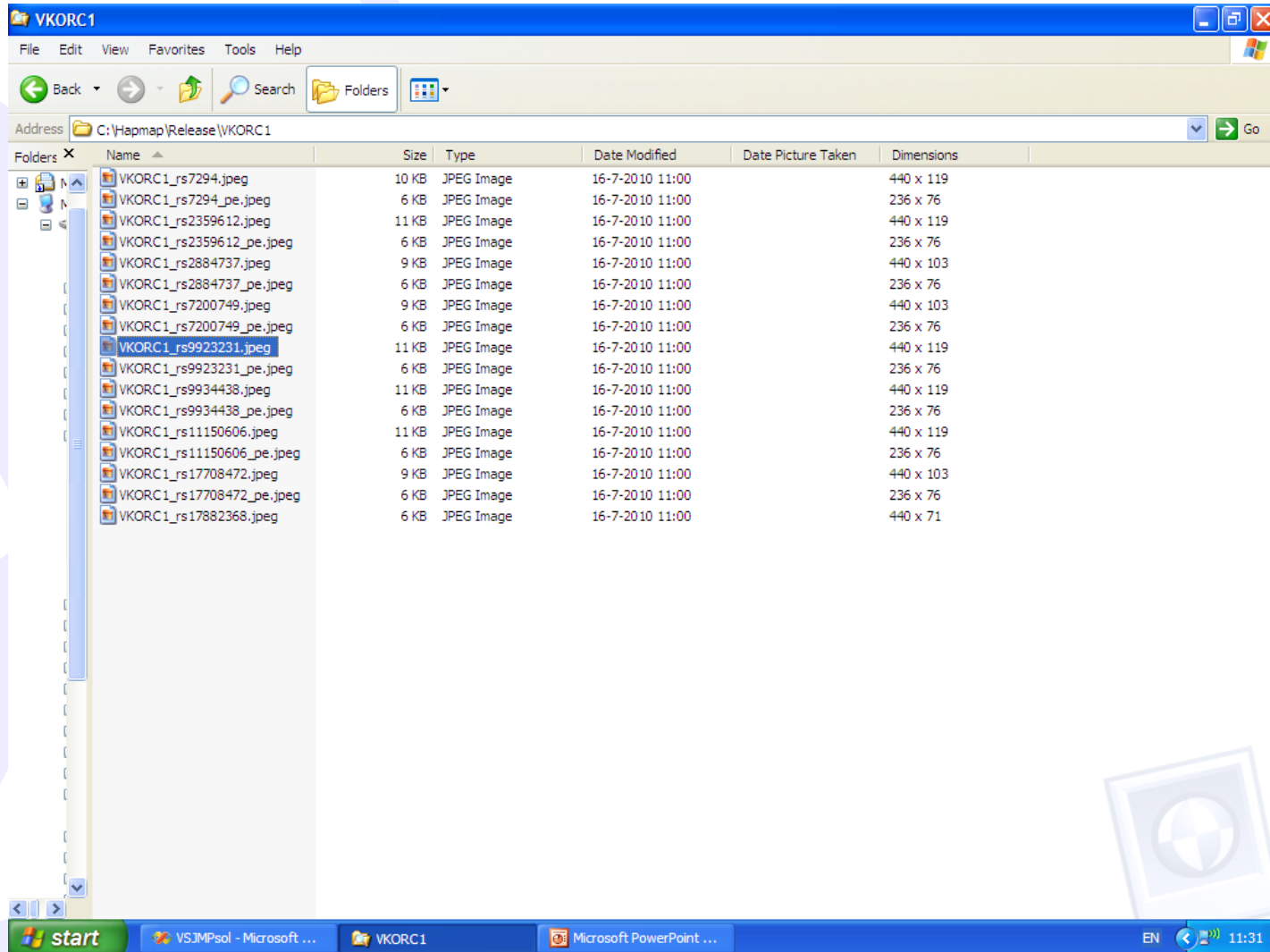
Results Restriction using Chi-Square Test

The screenshot displays the JMP software interface. On the left, the **JMPInAuto** workflow editor shows a process starting with 'Start Workflow', followed by a decision 'FTP or JMP', then 'FTP/Unzip', 'JMPScript1', 'Run JMPScript2', and finally 'Next or Close' leading to 'Close JMP'. A 'Continue' button is at the top. A 'Select Graphs' dialog box is open, showing 'All' and 'ChiSq p < 0.05' options.

On the right, the **JMP** window (OCS CONSULTING - TRIAL INST.) - VKORC1.com shows a data table with the following columns: rs#, chrom, refhom-gt, refhom-count, het-gt, het-count, and otherh. The table contains 21 rows of data for various SNPs on chromosome 16.

	rs#	chrom	refhom-gt	refhom-count	het-gt	het-count	otherh
5	rs2359612	chr16	A/A	3	A/G	15	G/G
6	rs9934438	chr16	G/G	48	A/G	6	A/A
7	rs13336384	chr16	G/G	54	A/G	3	A/A
8	rs17708472	chr16	G/G	51	A/G	6	A/A
9	rs2884737	chr16	A/A	52	A/C	2	C/C
10	rs9923231	chr16	C/C	48	C/T	6	T/T
11	rs11150606	chr16	T/T	106	C/T	7	C/C
12	rs7294	chr16	C/C	45	C/T	54	T/T
13	rs7200749	chr16	G/G	110	A/G	2	A/A
14	rs2359612	chr16	A/A	22	A/G	46	G/G
15	rs9934438	chr16	G/G	45	A/G	46	A/A
16	rs13336384	chr16	G/G	109	A/G	1	A/A
17	rs17708472	chr16	G/G	65	A/G	43	A/A
18	rs2884737	chr16	A/A	65	A/C	38	C/C
19	rs9923231	chr16	C/C	45	C/T	46	T/T
20	rs11150606	chr16	T/T	6	C/T	35	C/C
21	rs7294	chr16	C/C	89	C/T	18	T/T

The taskbar at the bottom shows the Windows start button, taskbar icons for Phuse2010, VSJMPsol (Runnin...), JMPInAuto, Select Graphs, and JMP (OCS CONSUL...), along with system icons and the time 23:16.

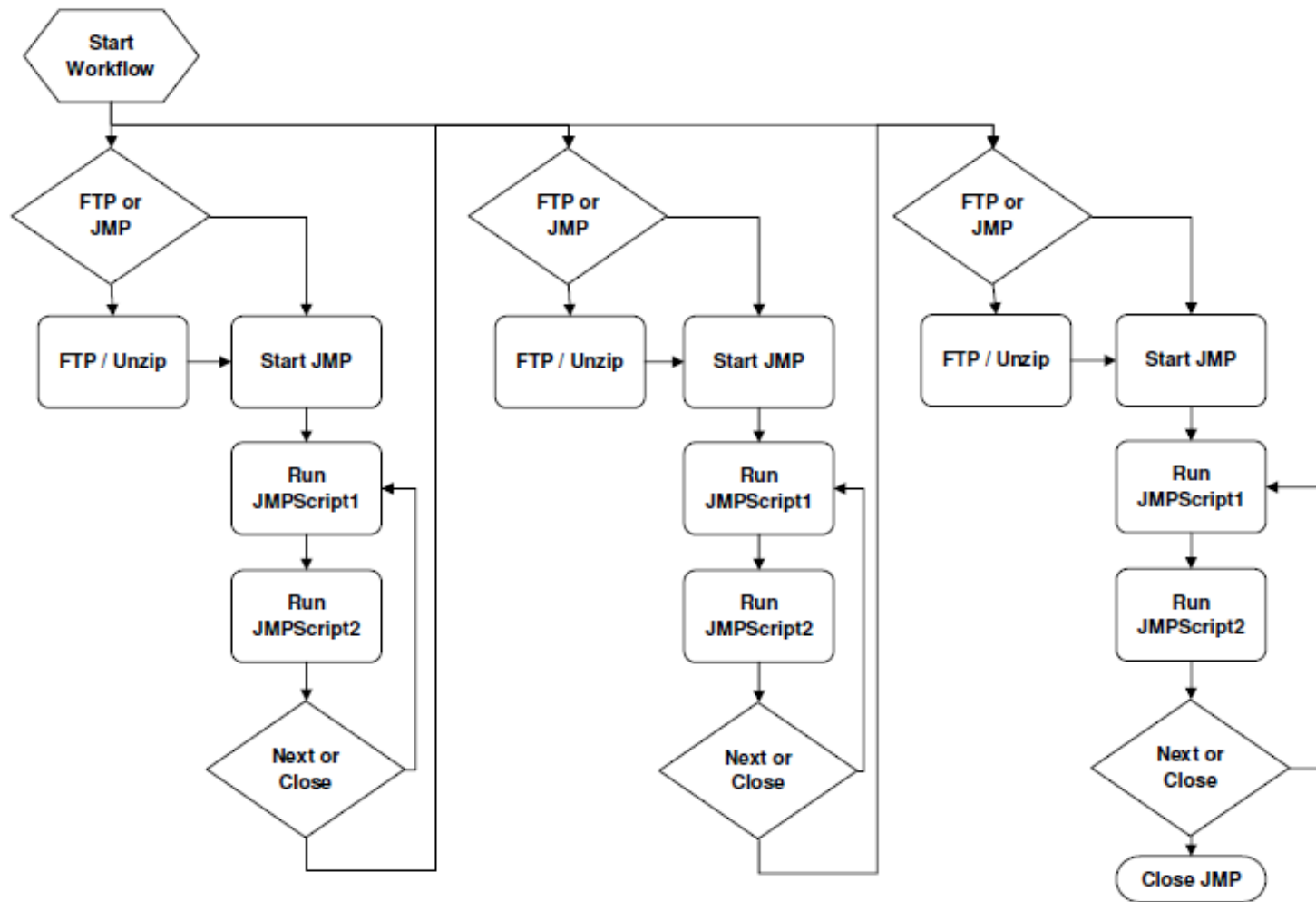


Windows Explorer window titled "VKORC1" showing a directory of JPEG images. The address bar indicates the path: C:\Hapmap\Release\VKORC1.

Name	Size	Type	Date Modified	Date Picture Taken	Dimensions
VKORC1_rs7294.jpeg	10 KB	JPEG Image	16-7-2010 11:00		440 x 119
VKORC1_rs7294_pe.jpeg	6 KB	JPEG Image	16-7-2010 11:00		236 x 76
VKORC1_rs2359612.jpeg	11 KB	JPEG Image	16-7-2010 11:00		440 x 119
VKORC1_rs2359612_pe.jpeg	6 KB	JPEG Image	16-7-2010 11:00		236 x 76
VKORC1_rs2884737.jpeg	9 KB	JPEG Image	16-7-2010 11:00		440 x 103
VKORC1_rs2884737_pe.jpeg	6 KB	JPEG Image	16-7-2010 11:00		236 x 76
VKORC1_rs7200749.jpeg	9 KB	JPEG Image	16-7-2010 11:00		440 x 103
VKORC1_rs7200749_pe.jpeg	6 KB	JPEG Image	16-7-2010 11:00		236 x 76
VKORC1_rs9923231.jpeg	11 KB	JPEG Image	16-7-2010 11:00		440 x 119
VKORC1_rs9923231_pe.jpeg	6 KB	JPEG Image	16-7-2010 11:00		236 x 76
VKORC1_rs9934438.jpeg	11 KB	JPEG Image	16-7-2010 11:00		440 x 119
VKORC1_rs9934438_pe.jpeg	6 KB	JPEG Image	16-7-2010 11:00		236 x 76
VKORC1_rs11150606.jpeg	11 KB	JPEG Image	16-7-2010 11:00		440 x 119
VKORC1_rs11150606_pe.jpeg	6 KB	JPEG Image	16-7-2010 11:00		236 x 76
VKORC1_rs17708472.jpeg	9 KB	JPEG Image	16-7-2010 11:00		440 x 103
VKORC1_rs17708472_pe.jpeg	6 KB	JPEG Image	16-7-2010 11:00		236 x 76
VKORC1_rs17882368.jpeg	6 KB	JPEG Image	16-7-2010 11:00		440 x 71

The taskbar at the bottom shows the following applications: start, VSJMPsol - Microsoft ..., VKORC1, and Microsoft PowerPoint ...

Different Model factors
Different Modelling Techniques
Taverna Workflow
Warfarin History
WWF Custom Activity
Tracking
Persistence
Error Handling
Workflow Re-use / Validation



VKORC1_rs9923231



Ethnic

VKORC1_rs9923231

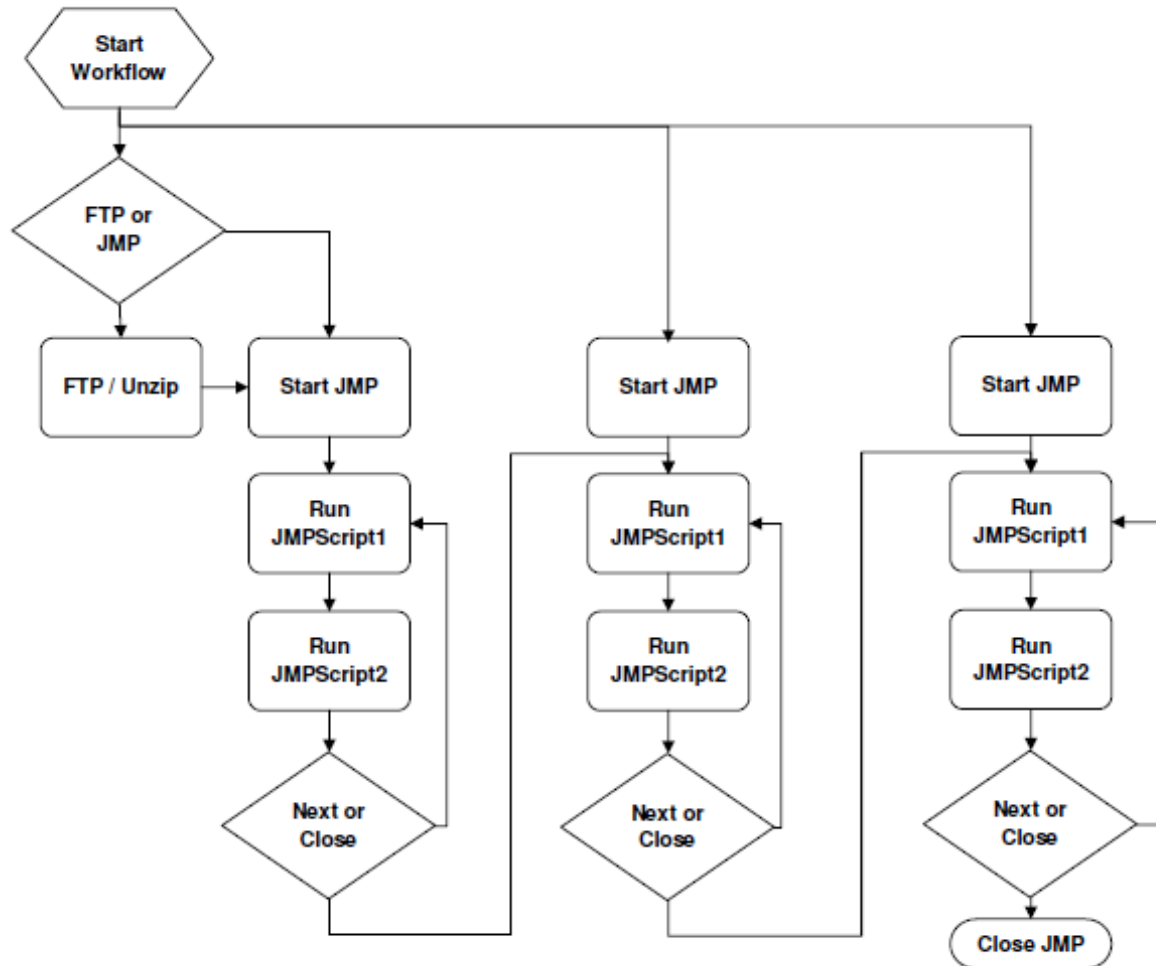


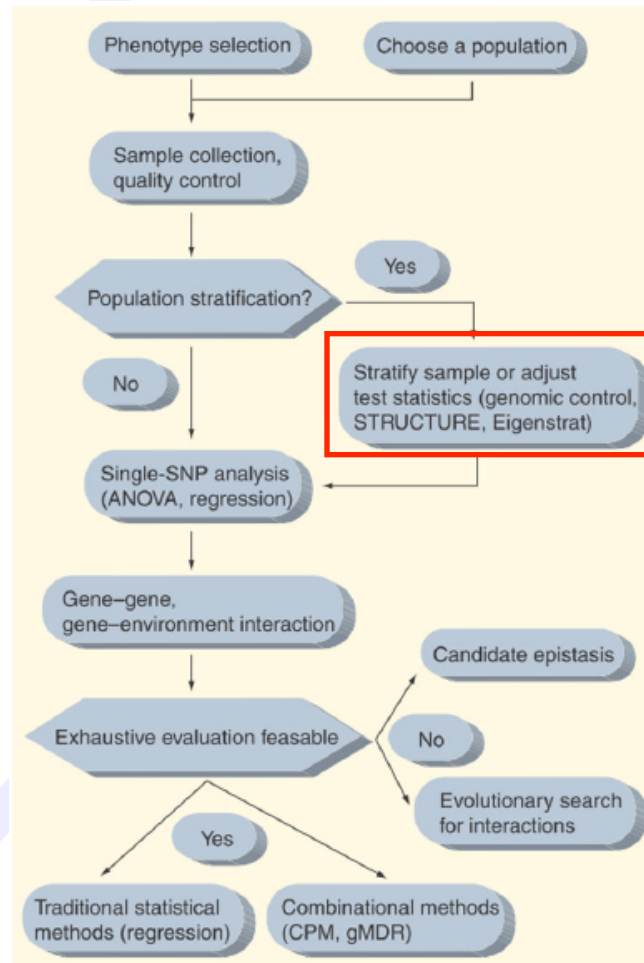
Clinical

VKORC1_rs9923231



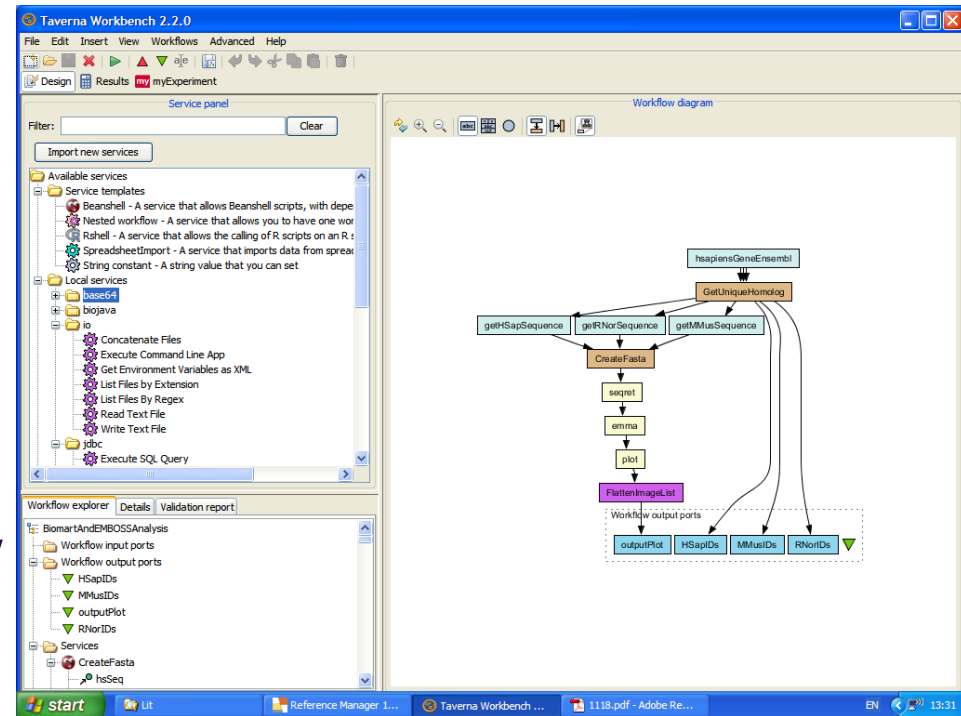
Pharma





From Turner SD, Crawford DC, Ritchie MD. Expert Rev Clin Pharmacol 2009;2(5):559-570

**Open Source
Beanshell Scripts
Rshell Scripts
(On – off) Data Flow
“UI”
Command Line App**



From <http://taverna.sourceforge.net>

Anti-clotting Agent

- → high responders
- → low responders

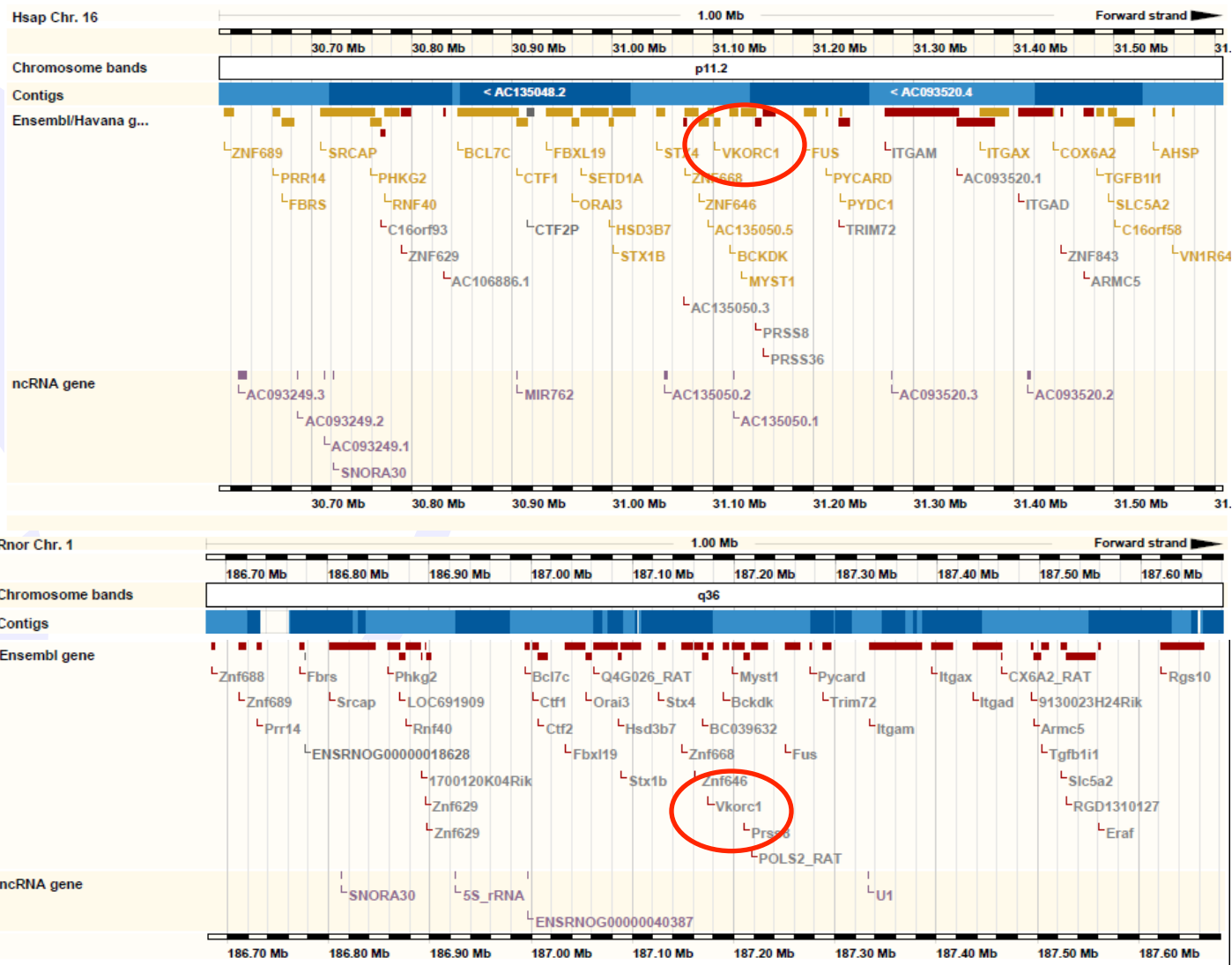
Rodenticide

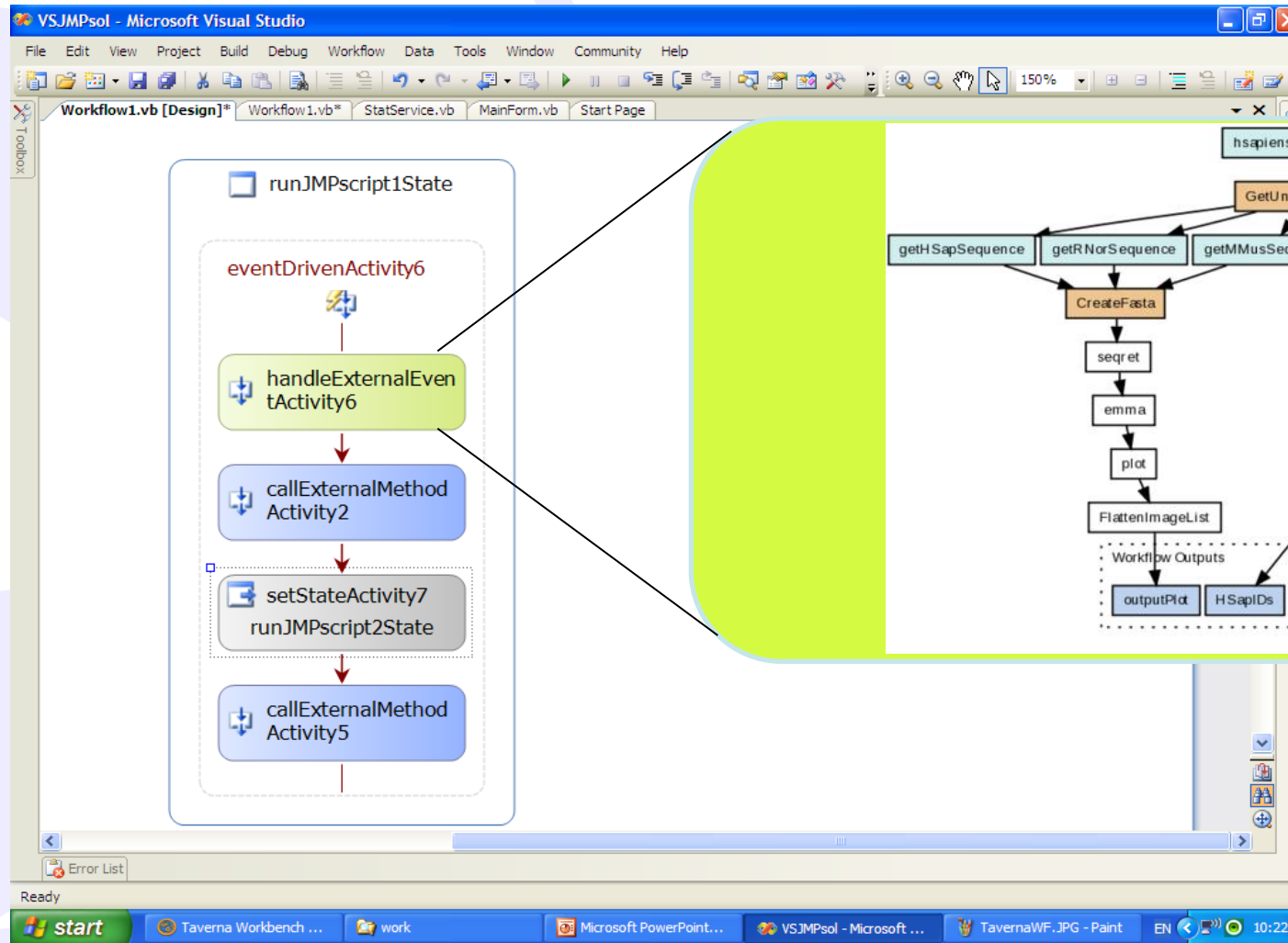
- → resistance

Rare Familial Diseases

- → coagulation factor deficiencies

Compare Genes of Species or Families





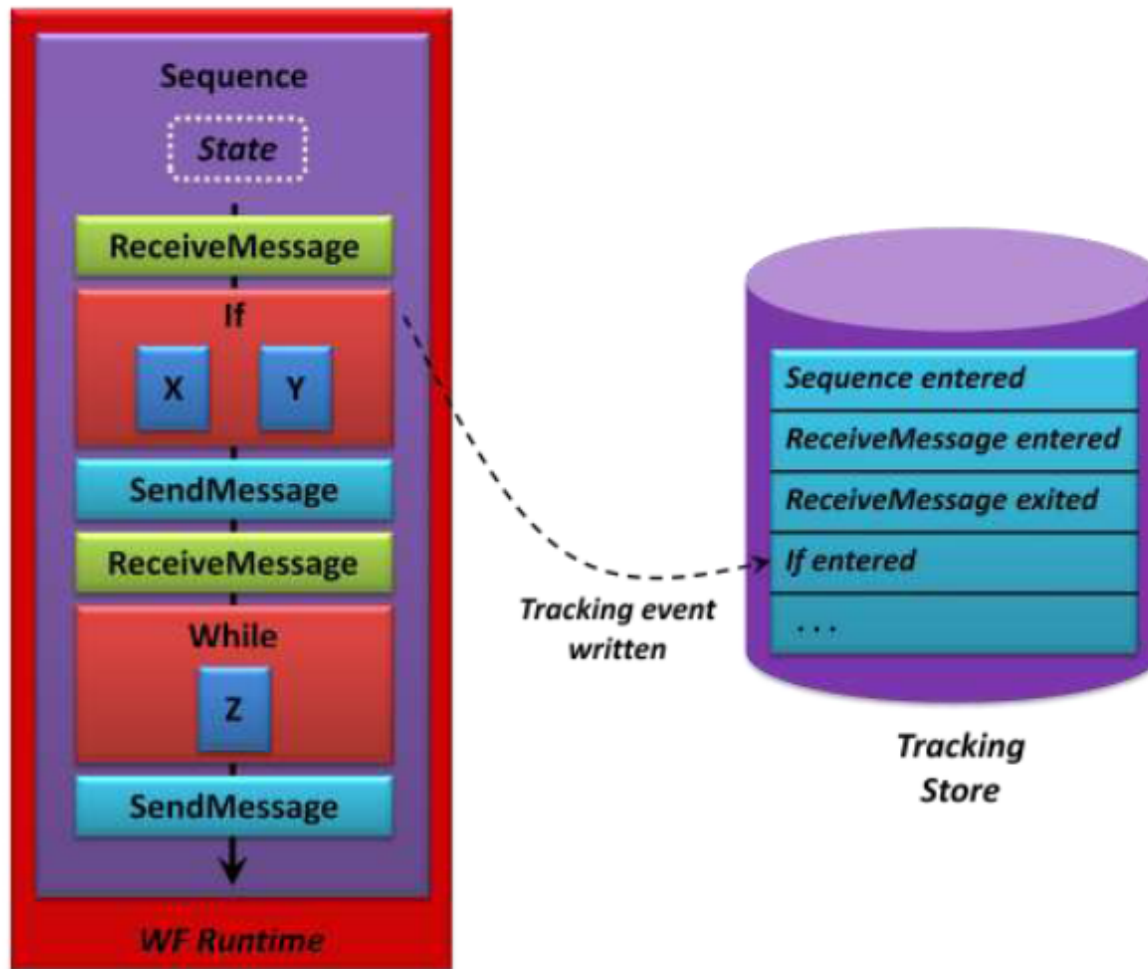
State 1

State 2

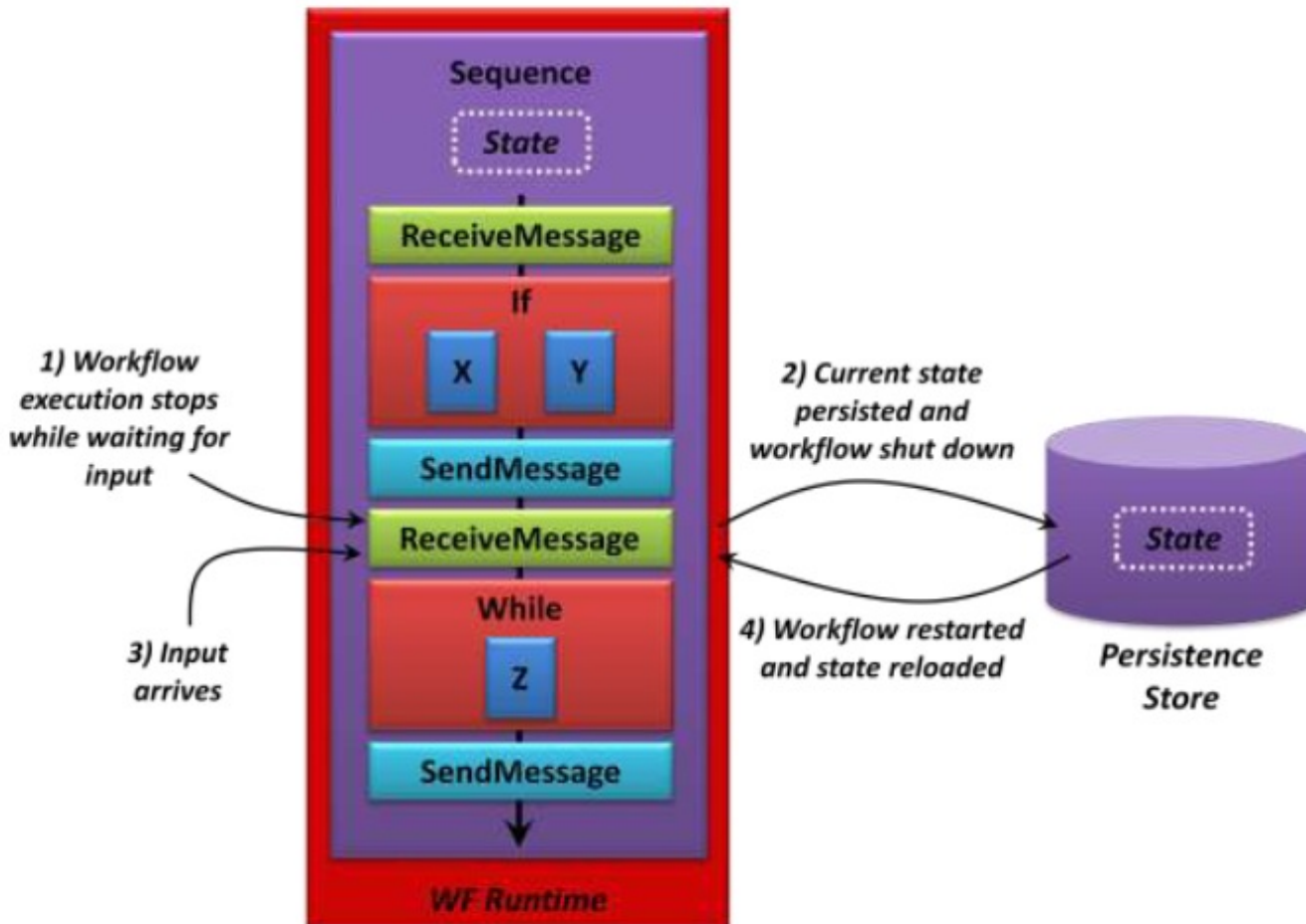
State 3

State 4

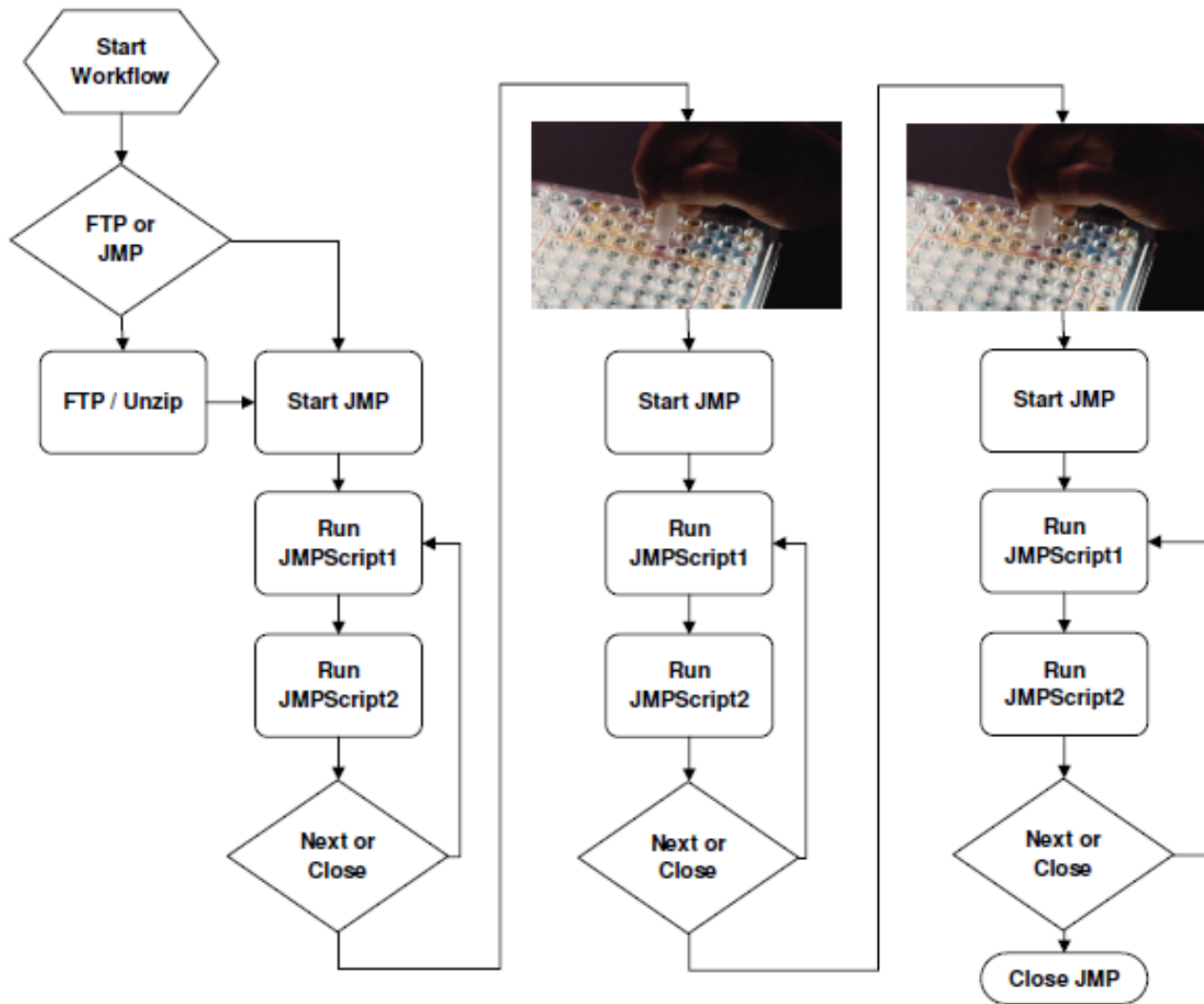
...

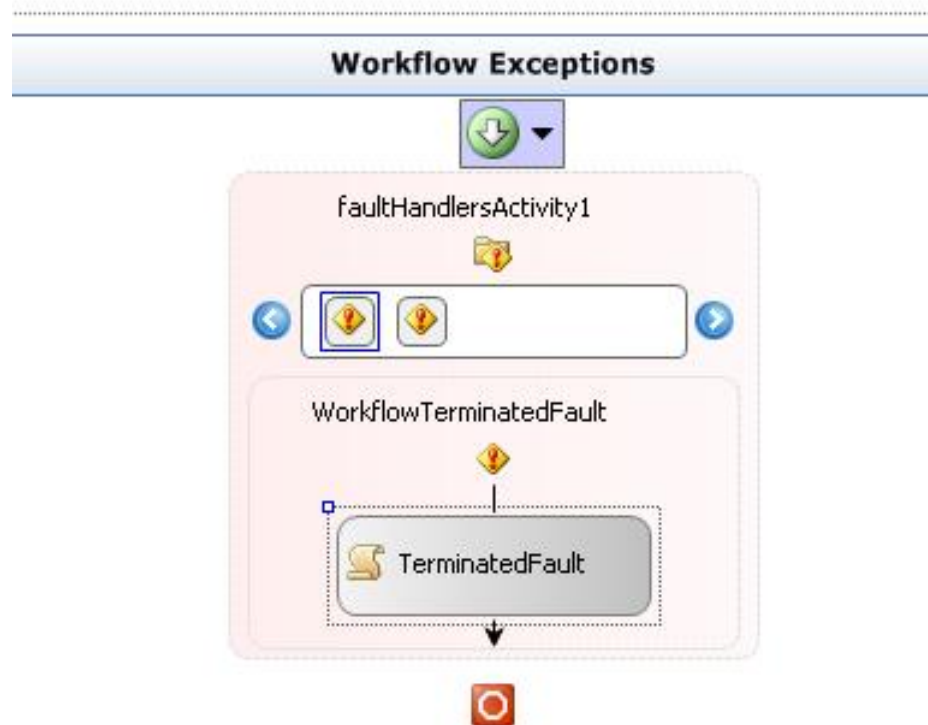


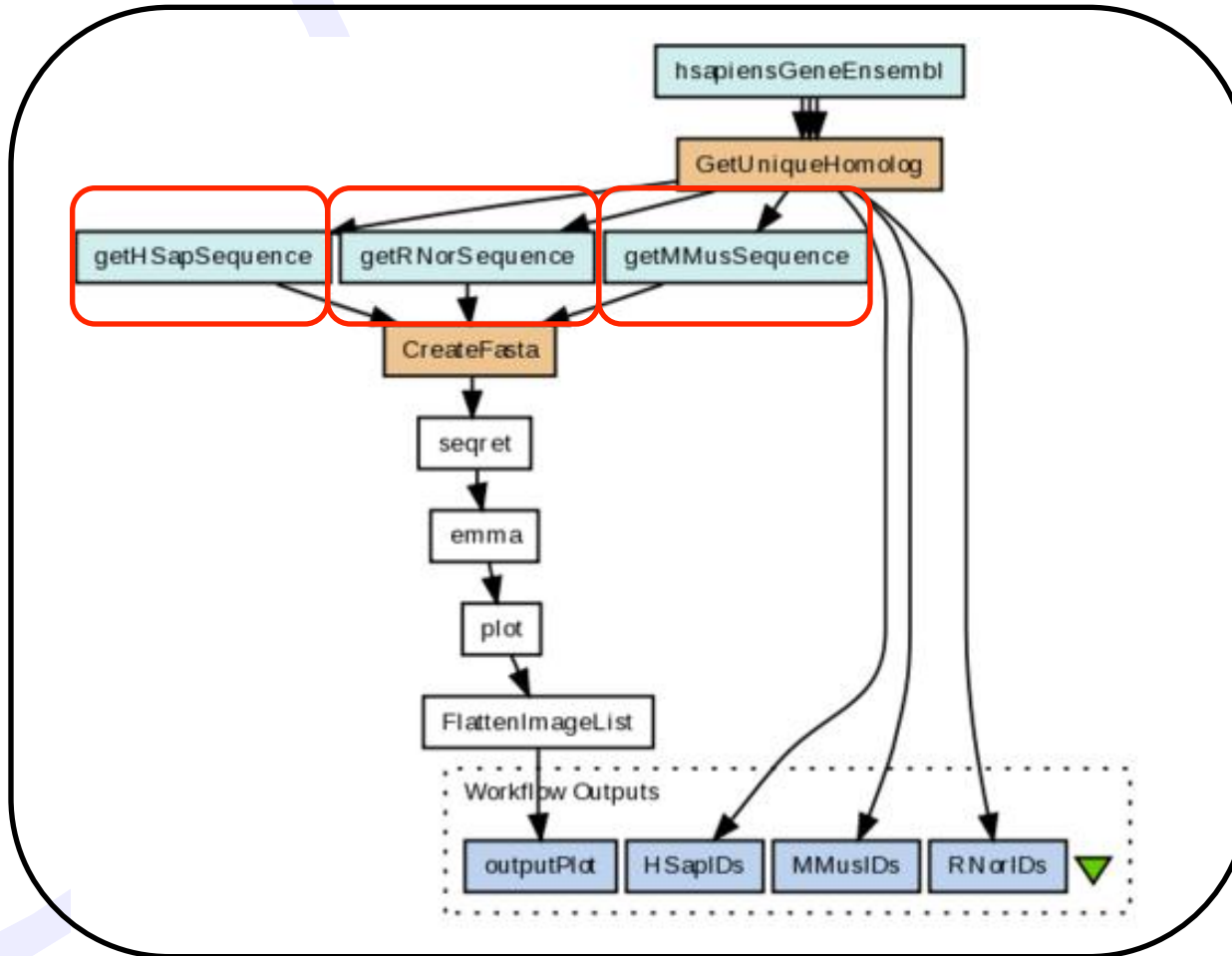
From Chappell D, The Workflow Way. Microsoft. 2009; 1-27.



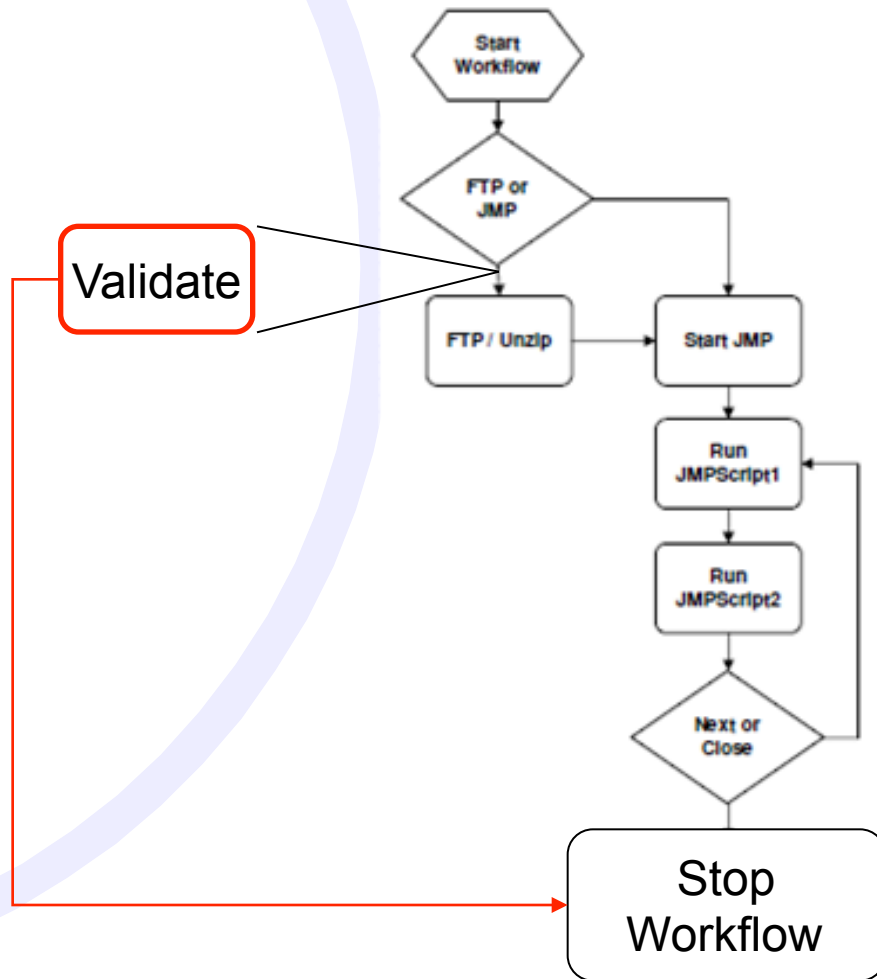
From Chappell D, The Workflow Way. Microsoft. 2009; 1-27.

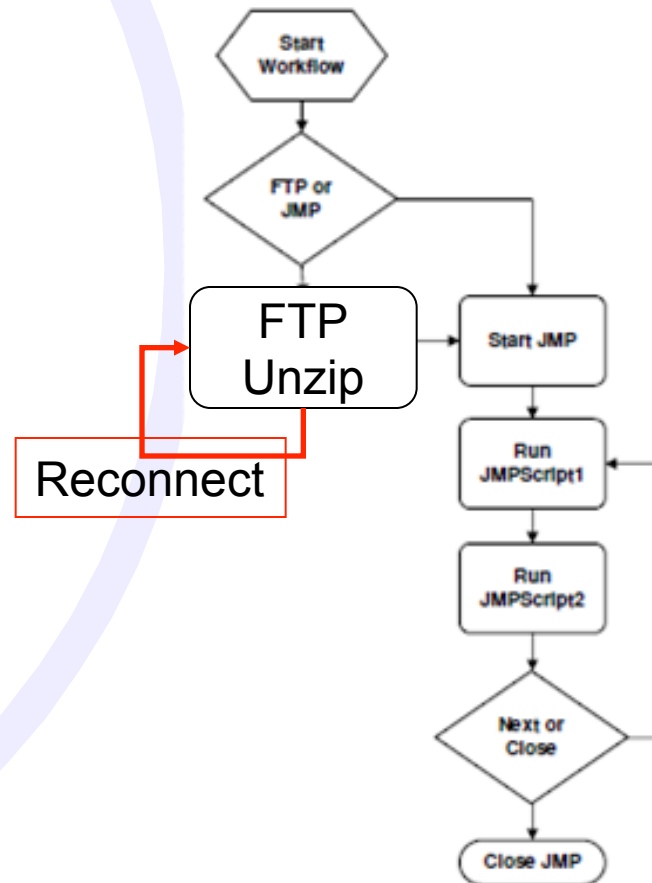


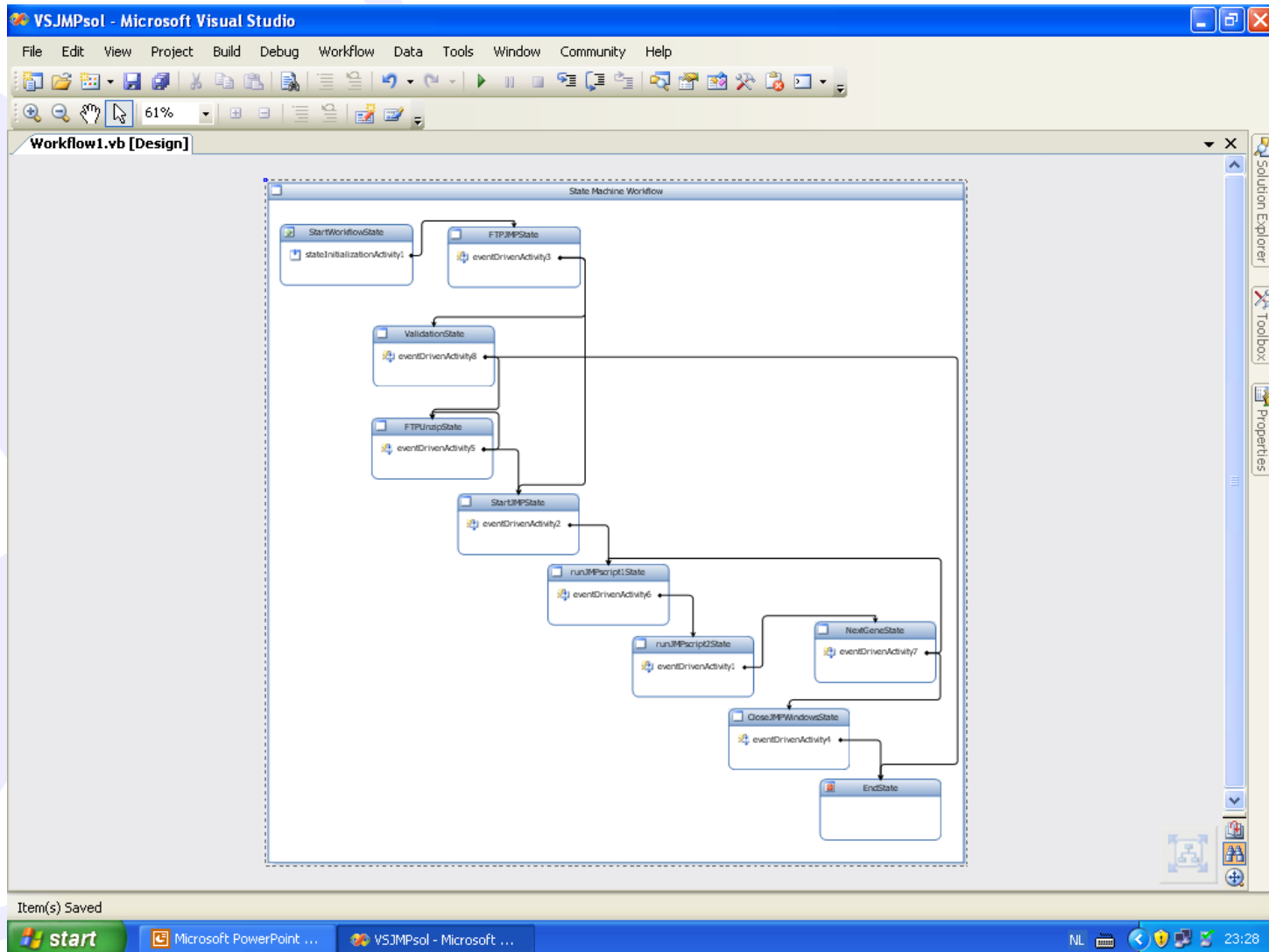




From <http://taverna.sourceforge.net>







It is possible to combine a data fetch and JMP scripts in a Windows workflow.

The graphical designer is attractive but not a Lego-like activity.

With more elaborate modelling or EDA workflows, the Tracking, Persistence and Error handling features probably make it worth considering.

?