

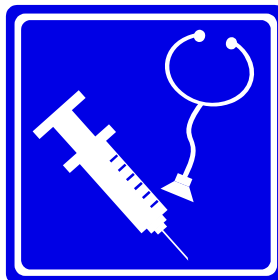


HNATyping User's Guide

Typing HNA types from whole-genome sequencing

Hsueh-Ting Chu
Htchu.taiwan@gmail.com

July, 2012



USER'S MANUAL

TABLE OF CONTENTS

Page 1-8

<i>Step 1: Prepare Whole Genome Sequencing Dataset</i>	1
1.1 Check the dataset ERX009608 at the European Nucleotide Archive	1
1.2 Download all the runs in the dataset through FTP	2
1.3 Download the HNATyping package from Sourceforge.net	2
1.4 Check the HNATyping package: (3 files)	2
<i>Step2: Filtering Reads by WGSReadFilter</i>	3
2.1 Open the template file:	3
2.2 Open the program: WGSReadFilter.exe	3
2.3 Paste the DNA templates	3
2.4 Select the directory for the entire WGS dataset	4
2.5 Set the output file	4
<i>Step3: Genotyping by WgsHnaTyping</i>	5
3.1 Open the filtered reads file	5
3.2 Typing result	5
3.3 Check the output files	5
3.4 Open the allele coverage file (.TXT)	6
3.5 Open the read alignment file (.ACE)	7
3.6 Check the allele loci	7

Revision Sheet

Release No.	Date	Revision Description
Rev. 0	2012/08/04	User's Guide

STEP 1: PREPARE WHOLE GENOME SEQUENCING DATASET

1.1 Check the dataset ERX009608 at the European Nucleotide Archive

The link of the dataset

Illumina paired-end WGS dataset

SRA Experiment: ERX009608 : Illumina Genome Analyzer II paired end sequencing

View: XML

Submitting Centre	Platform	Model	Read Count	Bas
Illumina Cambridge Ltd.	ILLUMINA	Illumina Genome Analyzer II	660,192,779	1330
Library Layout	Library Strategy	Library Source	Library Selection	Libr
PAIRED	WGS	GENOMIC	RANDOM	CT3

Description
Paired end sequencing (SRR000001) - ENA ERX009608 - Illumina Genome Analyzer II paired end sequencing

Study	Sample	Run	Organism	Instrument Model	Library Layout	Run Read Count	Run Base Count	ftp	Aspera	Galaxy
ERP000459	SRS003565	ERR024139	Homo sapiens	Illumina Genome Analyzer II	PAIRED	26,826,505	5Gb	Fastq file#1	not installed	Fastq file#1
ERP000459	SRS003565	ERR024139	Homo sapiens	Illumina Genome Analyzer II	PAIRED	26,826,505	5Gb	Fastq file#2	not installed	Fastq file#2
ERP000459	SRS003565	ERR024140	Homo sapiens	Illumina Genome Analyzer II	PAIRED	27,157,395	5Gb	Fastq file#1	not installed	Fastq file#1
ERP000459	SRS003565	ERR024140	Homo sapiens	Illumina Genome Analyzer II	PAIRED	27,157,395	5Gb	Fastq file#2	not installed	Fastq file#2
ERP000459	SRS003565	ERR024141	Homo sapiens	Illumina Genome Analyzer II	PAIRED	27,371,447	5Gb	Fastq file#1	not installed	Fastq file#1
ERP000459	SRS003565	ERR024141	Homo sapiens	Illumina Genome Analyzer II	PAIRED	27,371,447	5Gb	Fastq file#2	not installed	Fastq file#2
ERP000459	SRS003565	ERR024142	Homo sapiens	Illumina Genome Analyzer II	PAIRED	27,624,257	5Gb	Fastq file#1	not installed	Fastq file#1
ERP000459	SRS003565	ERR024142	Homo sapiens	Illumina Genome Analyzer II	PAIRED	27,624,257	5Gb	Fastq file#2	not installed	Fastq file#2
ERP000459	SRS003565	ERR024143	Homo sapiens	Illumina Genome Analyzer II	PAIRED	27,010,899	5Gb	Fastq file#1	not installed	Fastq file#1
ERP000459	SRS003565	ERR024143	Homo sapiens	Illumina Genome Analyzer II	PAIRED	27,010,899	5Gb	Fastq file#2	not installed	Fastq file#2

[Next](#)

1.2 Download all the runs in the dataset through FTP

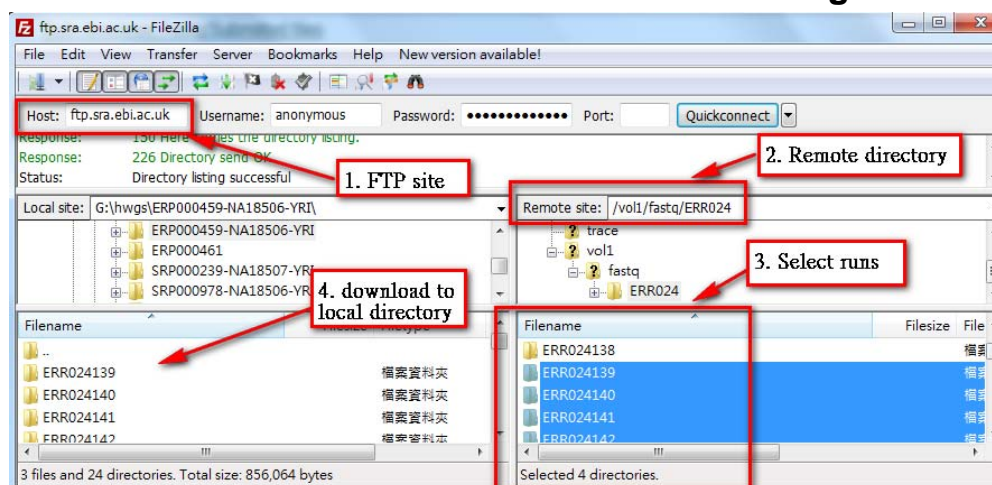
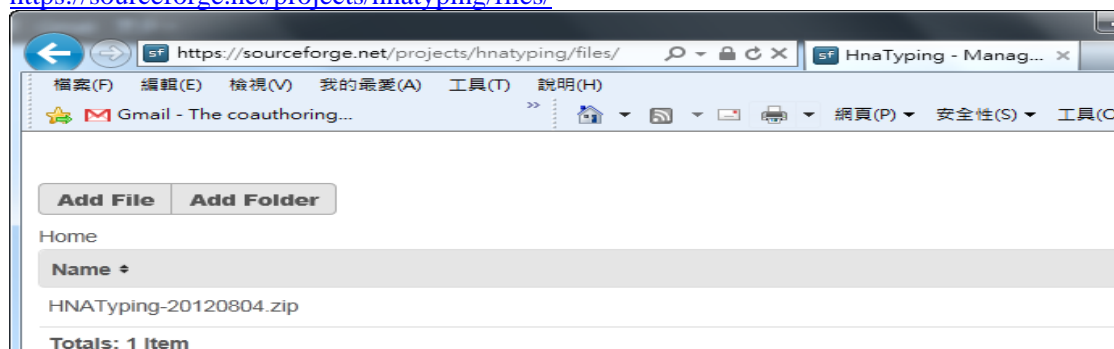


Table 1. List of runs in each Illumina WGS datasets

Samples	Study ERP	ERX	Runs
NA18507	ERP000460	ERX009609	ERR024163- ERR024186
NA18508	ERP000459	ERX009608	ERR024139- ERR024162
NA18506	ERP000461	ERX009610	ERR024201- ERR024200
NA12891	SRP000032	ERX000172	ERR001785- ERR001793
NA12892	SRP000032	ERX000174	ERR001827-ERR001868
NA12878	SRP000032	ERX000170	ERR001698-ERR002356

1.3 Download the HNATyping package from Sourceforge.net

<https://sourceforge.net/projects/hnatyping/files/>



1.4 Check the HNATyping package: (4 files)

HNA-DNA-Template.fa	2012/8/4 下午 05:22
HNATypingManual.pdf	2012/8/4 下午 07:17
WgsHnaTyping.exe	2012/8/4 下午 06:28
WGSReadFilter.exe	2012/8/4 下午 06:29

STEP2. FILTERING READS BY WGSREADFILTER

2.1 Open the template file:

```

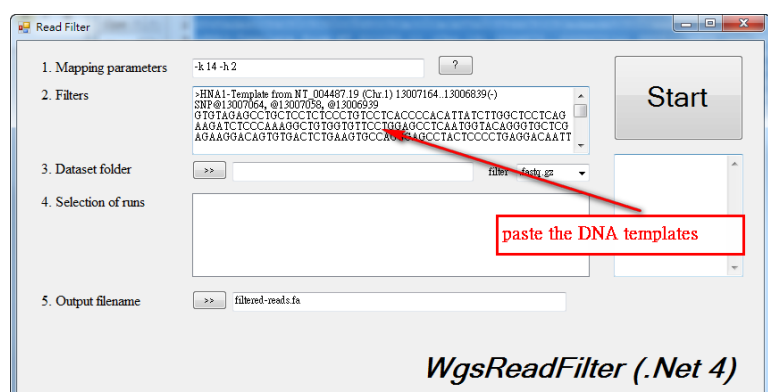
1 >HNA1-Template from NT_004487.19 (Chr.1) 13007164..13006839(-) SNP@13007064, @13007058, @13006939
2 GTGTAGAGCCTGCTCCTCTCCCTGTCCTCACCCACATTATCTTGGCTCCTCAGAAGATCTCCCAAAGGCTGTGGTGTTCCTGGAGCCTCAATGGTACAGGGTGTCTGAGAA
3 >HNA3-Template from NT_011295.11 (Chr.19) 2004872..2005072(+) SNP@2004972 (rs2288904)
4 TAGACTGTCCTGAGAGCACAGGTATGGGCAGTGGCAGTGTACTAGACTTGGGGACTCACCTCCCTCCATCTCTCTTGGCAGGGAGTGGCTGAGGTGCTTCAAGATGGTGACT
5 >HNA4-Template from NT_010393.16 (Chr.16) 31216711..31216911(+) SNP@31216811 (rs1143679/ rs17362505)
6 CACAGGGTGTGGTTGGAGCCCCCAGGAGATAGTGGCTGCCAACCAAGGGGCGAGCCTCTACCAGTGGGACTACAGCACAGGCTCATGCGAGCCCATCCGCTGCAGGGTG
7 >HNA5-Template from NT_010393.16 (Chr.16) 30457941..30458141(+) SNP@30458041 (rs59554592/ rs2230433)
8 TTGATTTATTTCTTTCTGGGCCACCAGATCCCTCAGTTCTGATATTTCCCAACCTGATCCTCCCTCCTCCATCTTTCCTGATCATCCCCACAGATCCAGAGCCCTGCGTC

```

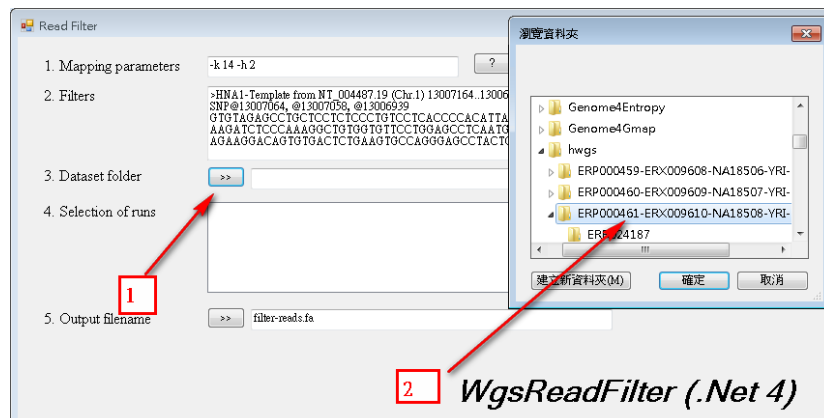
2.2 Open the program: WgsReadFilter.exe



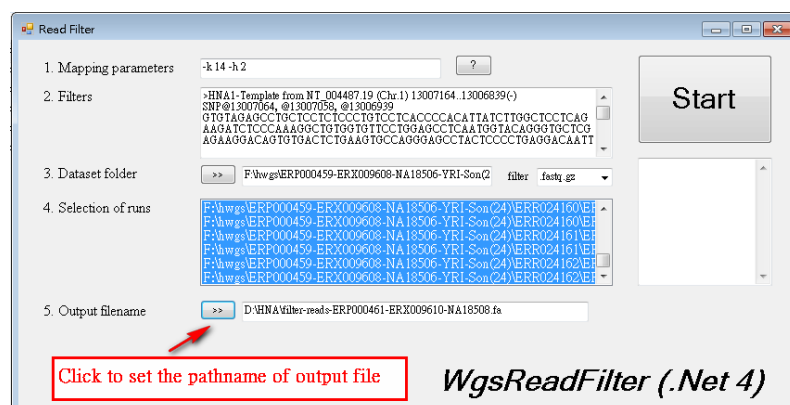
2.3 Paste the DNA templates



2.4 Select the directory for the entire WGS dataset

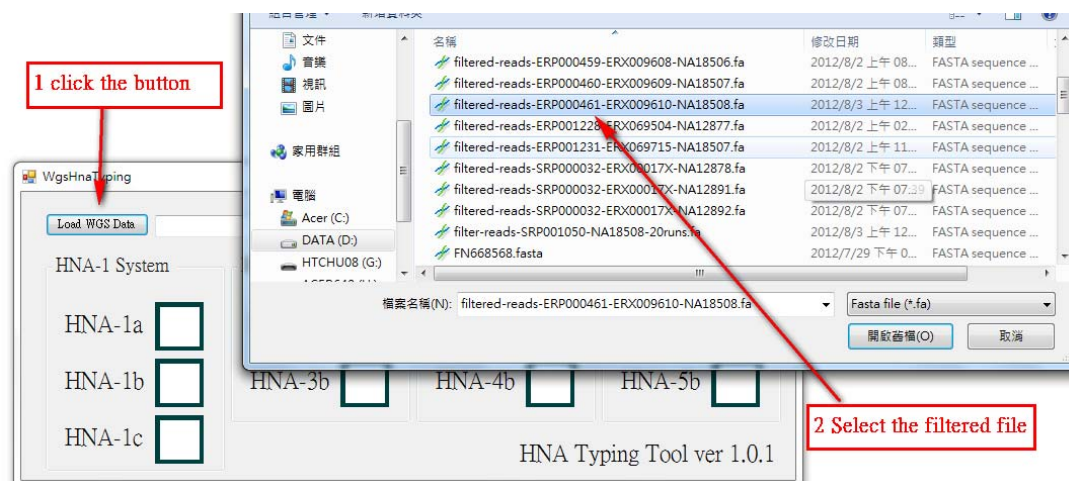


2.5 Set the output file

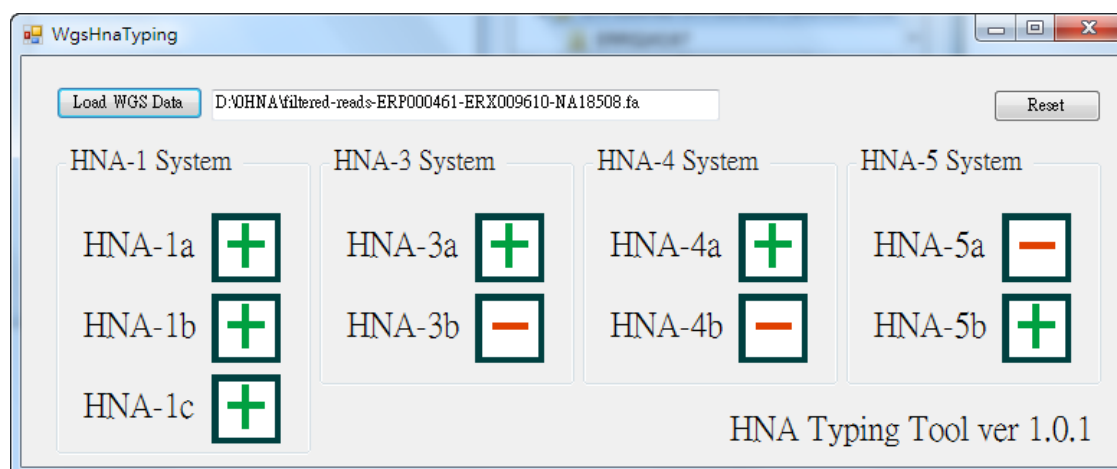


STEP3: GENOTYPING BY WGS HNATYPING

3.1 Open the filtered reads file



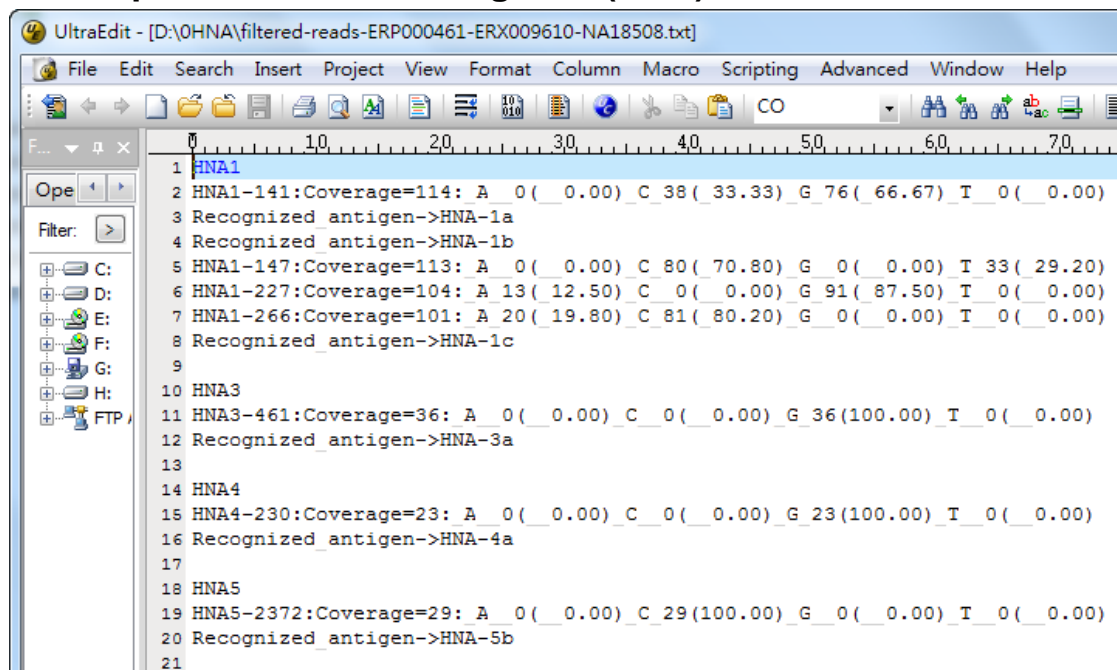
3.2 Typing result



3.3 Check the output files

	filtered-reads-ERP000461-ERX009610-NA18508.ace	2012/8/4 下午 07...	ACE 檔案
	filtered-reads-ERP000461-ERX009610-NA18508.txt	2012/8/4 下午 07...	TXT 檔案

3.4 Open the allele coverage file (.TXT)



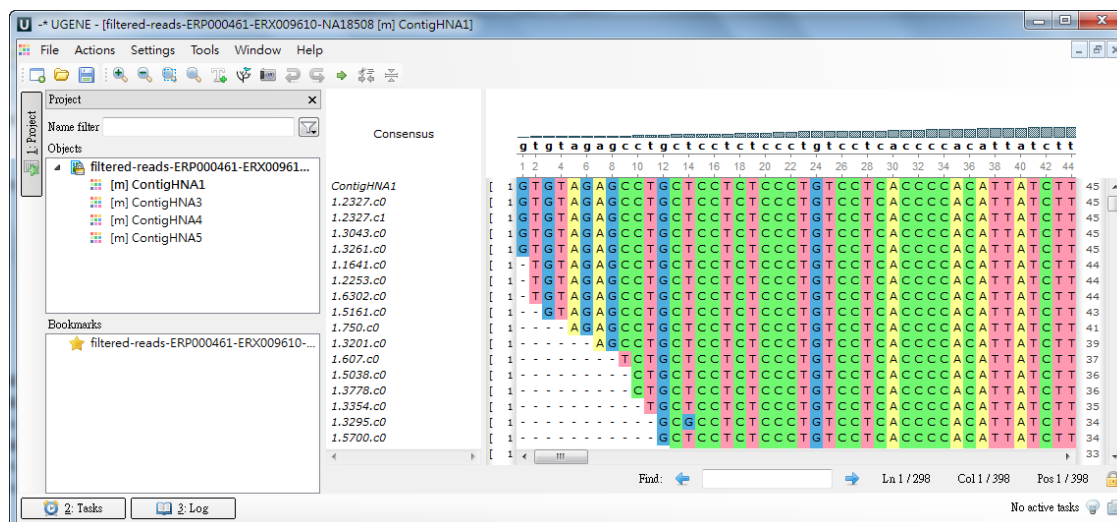
The screenshot shows the UltraEdit text editor with a file named "D:\0HNA\filtered-reads-ERP000461-ERX009610-NA18508.txt". The editor displays a list of HNA alleles and their coverage data. The data is organized into groups for HNA1, HNA3, HNA4, and HNA5, each followed by a "Recognized antigen" entry.

```

1 HNA1
2 HNA1-141:Coverage=114: A 0 ( 0.00) C 38 ( 33.33) G 76 ( 66.67) T 0 ( 0.00)
3 Recognized_antigen->HNA-1a
4 Recognized_antigen->HNA-1b
5 HNA1-147:Coverage=113: A 0 ( 0.00) C 80 ( 70.80) G 0 ( 0.00) T 33 ( 29.20)
6 HNA1-227:Coverage=104: A 13 ( 12.50) C 0 ( 0.00) G 91 ( 87.50) T 0 ( 0.00)
7 HNA1-266:Coverage=101: A 20 ( 19.80) C 81 ( 80.20) G 0 ( 0.00) T 0 ( 0.00)
8 Recognized_antigen->HNA-1c
9
10 HNA3
11 HNA3-461:Coverage=36: A 0 ( 0.00) C 0 ( 0.00) G 36 (100.00) T 0 ( 0.00)
12 Recognized_antigen->HNA-3a
13
14 HNA4
15 HNA4-230:Coverage=23: A 0 ( 0.00) C 0 ( 0.00) G 23 (100.00) T 0 ( 0.00)
16 Recognized_antigen->HNA-4a
17
18 HNA5
19 HNA5-2372:Coverage=29: A 0 ( 0.00) C 29 (100.00) G 0 ( 0.00) T 0 ( 0.00)
20 Recognized_antigen->HNA-5b
21

```

3.5 Open the read alignment file (.ACE)



3.6 Check the allele loci

