Jacek Radajewski
Student number: 43612772

SCIE2100 Prac 6

Question 1

Part a

Log transformed ratios are useful because they normalise the data. If we look at the samples obtained below we can see raw values, ratio and $\log_2$ ratio. One problem with ratio is that when Healthy probe is greater than that of ER Positive probe the ration will be between 0 and 1. If the reverse is true then the ratio values is between 1 and positive infinity. This problematic as the ratio values will be very skewed (see last row). Log ratio on the other hand will provide a close to normal distribution.

Healthy	ER positive	Ratio	$\log_2$ ratio
9510.4	11803.9	1.24115704913	0.311685677643
1804.1	2126.2	1.17853777507	0.236998001295
9.7	62.9	6.48453608247	2.69700336469

Code

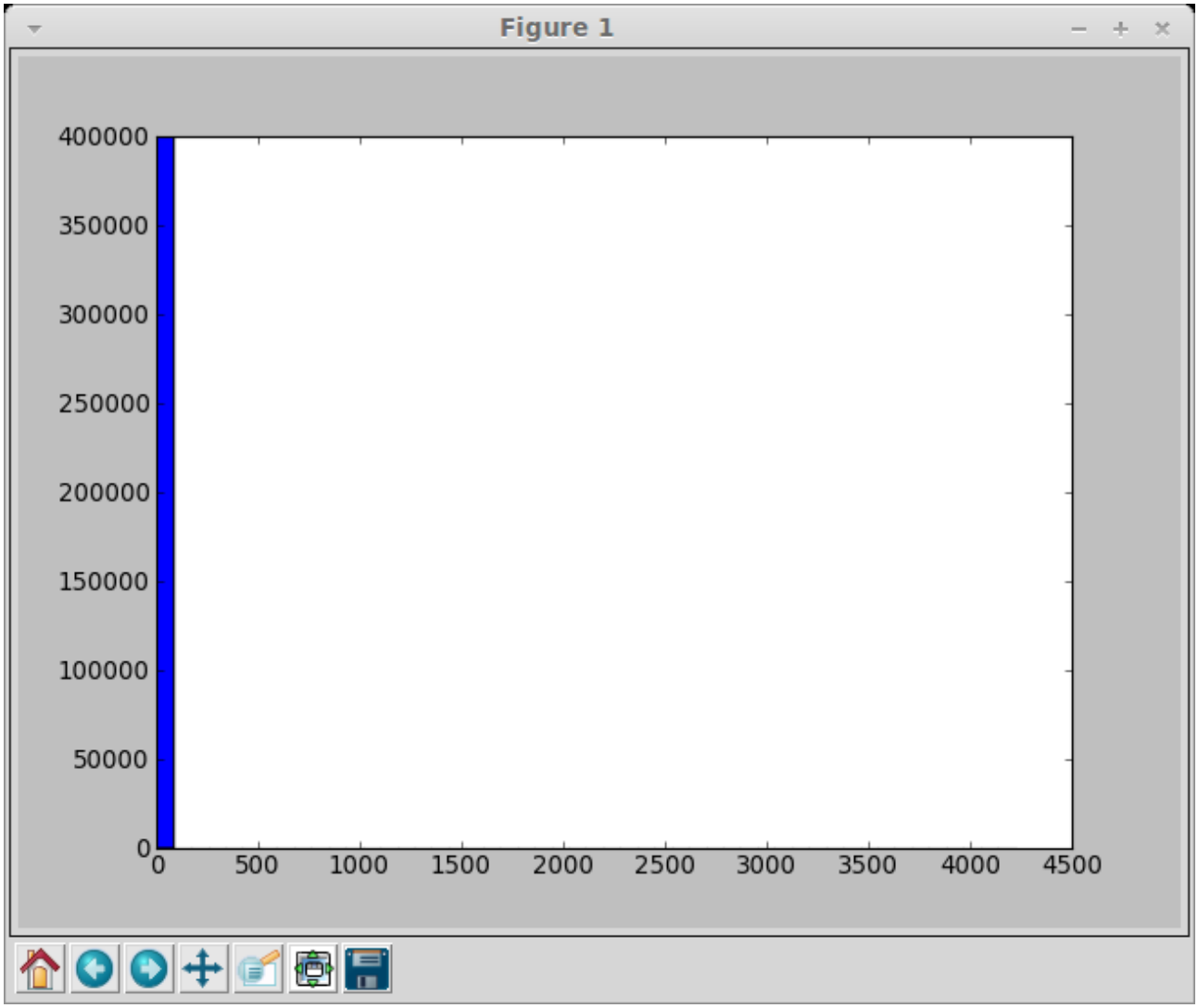
```
'''
Created on 07/05/2014

@author: s4361277
'''
from genome import *

ge3716 = readGEOFile("GDS3716.soft")

# get cancerous and healthe samples
# first ER+: GSM512557
# corresponding healthy: GSM512539
samples = ge3716.getSamples([ge3716.getHeaderIndex("GSM512557"), \
                             ge3716.getHeaderIndex("GSM512539")])

# create a list of probes
probes = []
probes.append(samples.get("200034_s_at"))
probes.append(samples.get("200599_s_at"))
probes.append(samples.get("211300_s_at"))

print "healthy\terplus\tratio\tlogratio"
for probe in probes:
    healthy = probe[1]
    erplus = probe[0]
    ratio = erplus / healthy
    logratio = math.log(ratio, 2)
    print healthy, erplus, ratio, logratio
```

Output

```
Data set GDS3716 contains 22215 genes
healthy   erplus      ratio logratio
9510.4 11803.9 1.24115704913 0.311685677643
1804.1 2126.2 1.17853777507 0.236998001295
9.7 62.9 6.48453608247 2.69700336469
```

Part b

Plot



Code

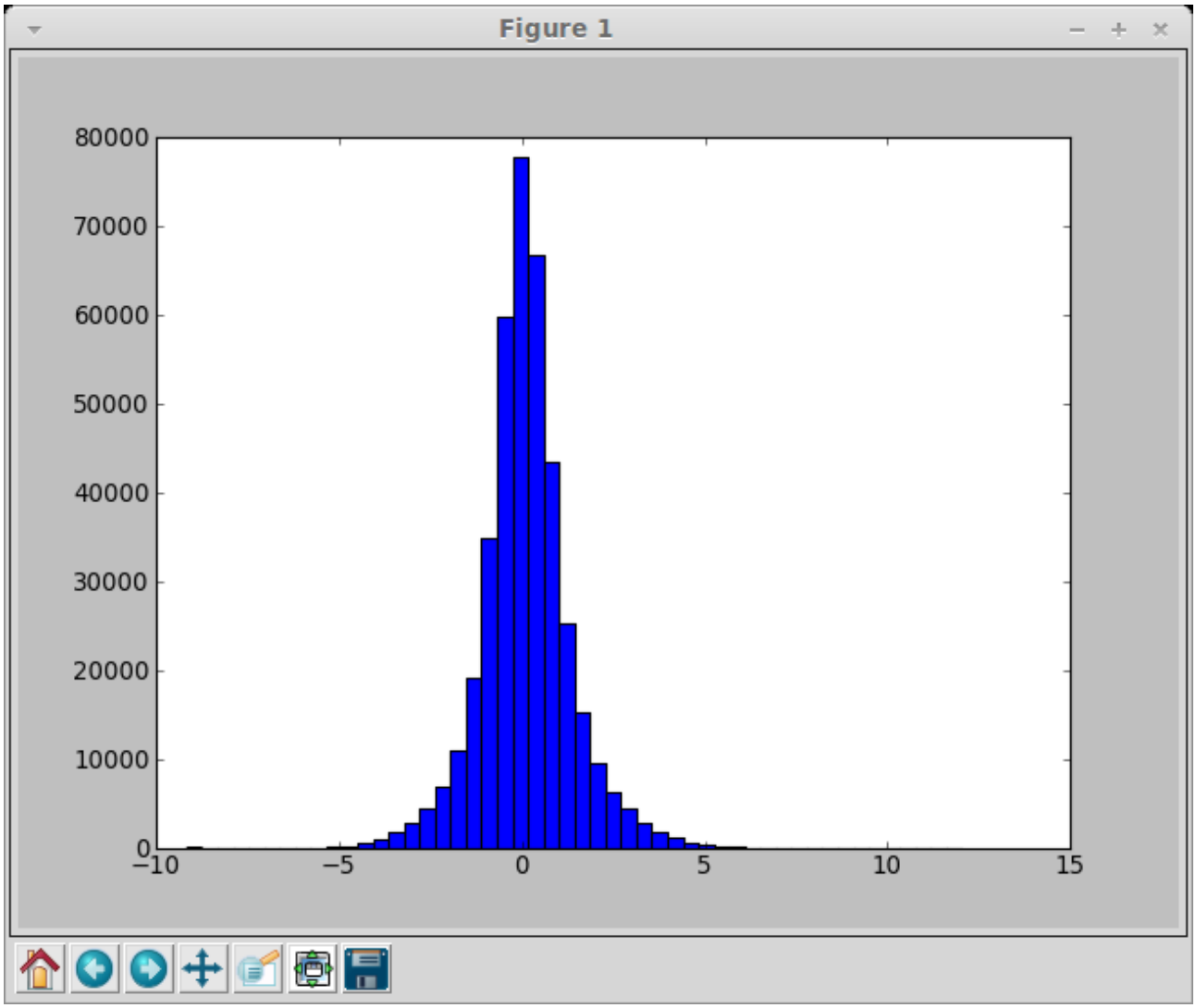
```
'''
Created on 07/05/2014

@author: s4361277
'''
```

```
from genome import *
import matplotlib.pyplot as plt
from sys import exit

ge3716 = readGEOFile("GDS3716.soft")

ratio = GeneExpression('GDS3716_ratio')

# set up age-matched sample indices (0: healthy, 1: cancer)
paired_ERpos = [[0, 1, 2, 3, 4, 5, 6, 7, 8], [33, 34, 35, 36, 37, 38, 39,
40, 41]]
paired_ERneg = [[9, 10, 11, 12, 13, 14, 15, 16, 17], [24, 25, 26, 27, 28,
29, 30, 31, 32]]


i = 0  # sample counter
while i < len(paired_ERpos[0]):
    name = 'S' + str(i + 1) + '_ER+/Healthy'  # meaningful name for column
header
    ratio.addSamples(name, ge3716.getRatio(paired_ERpos[1][i],
paired_ERpos[0][i]))
    i += 1

i = 0 # reset counter
while i < len(paired_ERneg[0]):
    name = 'S' + str(i + 1) + '_ER-/Healthy'  # meaningful name for column
header
    ratio.addSamples(name, ge3716.getRatio(paired_ERneg[1][i],
paired_ERneg[0][i]))
    i += 1

ratiovalues = ratio.matrix[:]
plt.hist(ratiovalues.ravel(), bins=50)
plt.show()
```

Part c

In contrast to the ratio histogram, the log ratio histogram is clearly (near) normally distributed

Plot



Code

```
'''
Created on 07/05/2014

@author: s4361277
'''
from genome import *
import matplotlib.pyplot as plt
from sys import exit

ge3716 = readGEOFile("GDS3716.soft")

logratio = GeneExpression('GDS3716_log_ratio')
```

```
# set up age-matched sample indices (0: healthy, 1: cancer)
paired_ERpos = [[0, 1, 2, 3, 4, 5, 6, 7, 8], \
                [33, 34, 35, 36, 37, 38, 39, 40, 41]]
paired_ERneg = [[9, 10, 11, 12, 13, 14, 15, 16, 17], \
                [24, 25, 26, 27, 28, 29, 30, 31, 32]]


i = 0  # sample counter
while i < len(paired_ERpos[0]):
    name = 'S' + str(i + 1) + '_ER+/Healthy'
    logratio.addSamples(name, ge3716.getLogRatio(paired_ERpos[1][i],\
                                                 paired_ERpos[0][i]))
    i += 1

i = 0 # reset counter
while i < len(paired_ERneg[0]):
    name = 'S' + str(i + 1) + '_ER-/Healthy'
    logratio.addSamples(name, ge3716.getLogRatio(paired_ERneg[1][i], \
                                                 paired_ERneg[0][i]))
    i += 1

logratiovalues = logratio.matrix[:]
plt.hist(logratiovalues.ravel(), bins=50)
plt.show()
```

Question 2

The following genes have been identified as significantly up- or down-regulated; breast cancer compared to healthy samples:

CFD
HLA-DQB1
INADL
CITED1
SCGB2A1
CXCL13
SOCS3
CCL2
CSN1S1
AW972855

Code

```
'''
Created on 07/05/2014

@author: s4361277
'''
from genome import *
import matplotlib.pyplot as plt

ge3716 = readGEOFile("GDS3716.soft", id_column = 1)

ratio = GeneExpression('GDS3716_ratio')
logratio = GeneExpression('GDS3716_log_ratio')


# set up age-matched sample indices (0: healthy, 1: cancer)
paired_ERpos = [[0, 1, 2, 3, 4, 5, 6, 7, 8], [33, 34, 35, 36, 37, 38, 39,
40, 41]]
paired_ERneg = [[9, 10, 11, 12, 13, 14, 15, 16, 17], [24, 25, 26, 27, 28,
29, 30, 31, 32]]

# Fill class with ER+ matched samples
i = 0  # sample counter

while i < len(paired_ERpos[0]):
    name = 'S' + str(i + 1) + '_ER+/Healthy'  # meaningful name for column
header
    ratio.addSamples(name, ge3716.getRatio(paired_ERpos[1][i],
paired_ERpos[0][i]))
    logratio.addSamples(name, ge3716.getLogRatio(paired_ERpos[1][i],
paired_ERpos[0][i]))
    i += 1

i = 0  # reset counter
while i < len(paired_ERneg[0]):
    name = 'S' + str(i + 1) + '_ER-/Healthy'  # meaningful name for column
header
    ratio.addSamples(name, ge3716.getRatio(paired_ERneg[1][i],
paired_ERneg[0][i]))
```

```
    logratio.addSamples(name, ge3716.getLogRatio(paired_ERneg[1][i],
paired_ERneg[0][i]))
    i += 1


sdict = {}  # dictionary for gene: no. of significant samples
# set up gene dictionary
genes = logratio.getGenes()
for gene in genes:  # fill sdict with 0 so can add counter to them
    sdict[gene] = 0

samples = []
samples.append(ge3716.getSamples([0, 33]))

# print samples

# Z-score calculation for each sample
for sample in range(0, 18):
    zscores = logratio.getZScore(sample)
    for key in zscores.iterkeys():
        # print key, zscores.get(key)
        if math.fabs(zscores.get(key)) > 2:
            sdict[key] += 1

print "List of all down or up regulated genes:"
for key in sdict:
    if sdict.get(key) > 10:
        print key, ":", sdict.get(key)
```

Output

```
Data set GDS3716 contains 14024 genes
List of all down or up regulated genes:
CFD : 11
HLA-DQB1 : 12
INADL : 11
CITED1 : 11
SCGB2A1 : 11
CXCL13 : 11
SOCS3 : 11
CCL2 : 12
CSN1S1 : 11
AW972855 : 11
```

Question 3

Part a

The top 5 genes whose expression patterns most strongly correlate with that of CLN2 are :

1. RNR1
2. MCD1
3. RAD27
4. AXL2
5. POL30

See code below for implementation.

Part b

From the gene ontology obtained below we can see that the following of the top 5 genes are known to be involved in the cell cycle:

1. partially RNR1: DNA replication.
2. MAIN: MCD1: mitosis, chromosome segregation
3. somewhat involved RAD27: DNA replication, removal of RNA primer
4. quite involved POL30: DNA replication, replication fork.

Code

```
'''
Created on 07/05/2014

@author: s4361277
'''
from genome_import *
import matplotlib.pyplot as plt
from webservice_import *

ge38 = readGEOFile('GDS38.soft', id_column=1)
cln2_genes = ge38.getGenes("CLN2")
# 1497 is the probe id obtained from the file
cln2R = ge38.getPearson("CLN2")

gecln2R = GeneExpression("CLN2", [1], cln2R)
gecln2Rsorted = gecln2R.sort(0)

# cut the top 5 entries into a new list
# cln2 will be in the first position so we exclude it
top5 = gecln2Rsorted[1:6]

for gene in top5:
    print "-------------", gene, "Gene Ontology term names: "
    rows = search("taxonomy:4932+AND+gene:" + gene, "uniprot",
format="list")
    terms = getGOTerms(rows[0])
    count = 0
    for term in terms:
        count += 1
        defs = getGODef(term)
```

```
        print count, ". ", defs.get("name")
```

Output

```
Data set GDS38 contains 6183 genes
Data set has 7499 null-values
------------- RNR1 Gene Ontology term names:
1 .  ribonucleoside-diphosphate reductase complex
2 .  cytoplasm
3 .  catalytic activity
4 .  nucleotide binding
5 .  ribonucleoside-diphosphate reductase activity, thioredoxin disulfide
as acceptor
6 .  deoxyribonucleotide biosynthetic process
7 .  DNA replication
8 .  oxidation-reduction process
9 .  metabolic process
10 .  ATP binding
11 .  oxidoreductase activity
------------- MCD1 Gene Ontology term names:
1 .  protein binding
2 .  mitosis
3 .  establishment of mitotic sister chromatid cohesion
4 .  DNA unwinding involved in DNA replication
5 .  nucleus
6 .  apoptotic process
7 .  mitochondrion
8 .  cellular response to DNA damage stimulus
9 .  protein acetylation
10 .  cell division
11 .  condensed nuclear chromosome
12 .  nuclear mitotic cohesin complex
13 .  mitotic chromosome condensation
14 .  chromosome segregation
15 .  nuclear chromosome
16 .  chromatin binding
17 .  chromosome
18 .  chromosome, centromeric region
19 .  cell cycle
20 .  double-strand break repair
------------- RAD27 Gene Ontology term names:
1 .  5'-flap endonuclease activity
2 .  nucleic acid phosphodiester bond hydrolysis
3 .  DNA binding
4 .  5'-3' exonuclease activity
5 .  nucleolus
6 .  exonuclease activity
7 .  magnesium ion binding
8 .  mitochondrion
9 .  nucleoplasm
10 .  DNA replication
11 .  metal ion binding
12 .  hydrolase activity
13 .  base-excision repair
14 .  DNA repair
15 .  hydrolase activity, acting on ester bonds
16 .  cellular response to DNA damage stimulus
```

```
17 .  catalytic activity
18 .  nuclease activity
19 .  endonuclease activity
20 .  DNA catabolic process, endonucleolytic
21 .  DNA replication, removal of RNA primer
22 .  nucleus
------------- AXL2 Gene Ontology term names:
1 .  septin ring
2 .  protein binding
3 .  integral component of membrane
4 .  membrane
5 .  molecular_function
6 .  calcium ion binding
7 .  integral component of plasma membrane
8 .  plasma membrane
9 .  cellular bud neck
10 .  axial cellular bud site selection
------------- POL30 Gene Ontology term names:
1 .  chromatin silencing at silent mating-type cassette
2 .  establishment of mitotic sister chromatid cohesion
3 .  maintenance of DNA trinucleotide repeats
4 .  mismatch repair
5 .  DNA binding
6 .  mitotic sister chromatid cohesion
7 .  chromatin silencing at telomere
8 .  postreplication repair
9 .  nucleus
10 .  replication fork
11 .  DNA replication
12 .  identical protein binding
13 .  DNA polymerase processivity factor activity
14 .  mitotic cell cycle
15 .  DNA repair
16 .  nucleotide-excision repair
17 .  cellular response to DNA damage stimulus
18 .  PCNA complex
19 .  lagging strand elongation
20 .  leading strand elongation
21 .  meiotic mismatch repair
22 .  regulation of DNA replication
23 .  error-free translesion synthesis
24 .  protein binding
25 .  positive regulation of exodeoxyribonuclease activity
```

Appendix A – Question 3b full search results

This is the full output including descriptions

```
Data set GDS38 contains 6183 genes
Data set has 7499 null-values
-------------------------------  RNR1  ----------------------
GO:0005971 ; ribonucleoside-diphosphate reductase complex ; An enzyme
complex composed of 2-4 or more subunits, which usually contains nonheme
iron and requires ATP for catalysis. Catalyzes the formation of
2'-deoxyribonucleoside diphosphate from ribonucleoside diphosphate, using
either thioredoxin disulfide or glutaredoxin disulfide as an acceptor.
GO:0005737 ; cytoplasm ; All of the contents of a cell excluding the plasma
membrane and nucleus, but including other subcellular structures.
GO:0003824 ; catalytic activity ; Catalysis of a biochemical reaction at
physiological temperatures. In biologically catalyzed reactions, the
reactants are known as substrates, and the catalysts are naturally
occurring macromolecular substances known as enzymes. Enzymes possess
specific binding sites for substrates, and are usually composed wholly or
largely of protein, but RNA that has catalytic activity (ribozyme) is often
also regarded as enzymatic.
GO:0000166 ; nucleotide binding ; Interacting selectively and
non-covalently with a nucleotide, any compound consisting of a nucleoside
that is esterified with (ortho)phosphate or an oligophosphate at any
hydroxyl group on the ribose or deoxyribose.
GO:0004748 ; ribonucleoside-diphosphate reductase activity, thioredoxin
disulfide as acceptor ; Catalysis of the reaction: 2'-deoxyribonucleoside
diph----------------------osphate + thioredoxin disulfide + H2O =
ribonucleoside diphosphate + thioredoxin. Thioredoxin disulfide is the
oxidized form of thioredoxin.
GO:0009263 ; deoxyribonucleotide biosynthetic process ; The chemical
reactions and pathways resulting in the formation of a deoxyribonucleotide,
a compound consisting of deoxyribonucleoside (a base linked to a
deoxyribose sugar) esterified with a phosphate group at either the 3' or
5'-hydroxyl group of the sugar.
GO:0006260 ; DNA replication ; The cellular metabolic process in which a
cell duplicates one or more molecules of DNA. DNA replication begins when
specific sequences, known as origins of replication, are recognized and
bound by initiation proteins, and ends when the original DNA molecule has
been completely duplicated and the copies topologically separated. The unit
of replication usually corresponds to the genome of the cell, an organelle,
or a virus. The template for replication can either be an existing DNA
molecule or RNA.
GO:0055114 ; oxidation-reduction process ; A metabolic process that results
in the removal or addition of one or more electrons to or from a substance,
with or without the concomitant removal or addition of a proton or protons.
GO:0008152 ; metabolic process ; The chemical reactions and pathways,
including anabolism and catabolism, by which living organisms transform
chemical substances. Metabolic processes typically transform small
molecules, but also include macromolecular processes such as DNA repair and
replication, and protein synthesis and degradation.
GO:0005524 ; ATP binding ; Interacting selectively and non-covalently with
ATP, adenosine 5'-triphosphate, a universally important coenzyme and enzyme
regulator.
GO:0016491 ; oxidoreductase activity ; Catalysis of an oxidation-reduction
(redox) reaction, a reversible chemical reaction in which the oxidation
state of an atom or atoms within a molecule is altered. One substrate acts
as a hydrogen or electron donor and becomes oxidized, while the other acts
as hydrogen or electron acceptor and becomes reduced.
-------------------------------  MCD1  ----------------------
GO:0005515 ; protein binding ; Interacting selectively and non-covalently
```

with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).
GO:0007067 ; mitosis ; A cell cycle process comprising the steps by which the nucleus of a eukaryotic cell divides; the process involves condensation of chromosomal DNA into a highly compacted form. Canonically, mitosis produces two daughter nuclei whose chromosome complement is identical to that of the mother cell.
GO:0034087 ; establishment of mitotic sister chromatid cohesion ; The process in which the sister chromatids of a replicated chromosome become joined along the entire length of the chromosome during S phase during a mitotic cell cycle.
GO:0006268 ; DNA unwinding involved in DNA replication ; The process in which interchain hydrogen bonds between two strands of DNA are broken or 'melted', generating unpaired template strands for DNA replication.
GO:0005634 ; nucleus ; A membrane-bounded organelle of eukaryotic cells in which chromosomes are housed and replicated. In most cells, the nucleus contains all of the cell's chromosomes except the organellar chromosomes, and is the site of RNA synthesis and processing. In some species, or in specialized cell types, RNA metabolism or DNA replication may be absent.
GO:0006915 ; apoptotic process ; A programmed cell death process which begins when a cell receives an internal (e.g. DNA damage) or external signal (e.g. an extracellular death ligand), and proceeds through a series of biochemical events (signaling pathway phase) which trigger an execution phase. The execution phase is the last step of an apoptotic process, and is typically characterized by rounding-up of the cell, retraction of pseudopodes, reduction of cellular volume (pyknosis), chromatin condensation, nuclear fragmentation (karyorrhexis), plasma membrane blebbing and fragmentation of the cell into apoptotic bodies. When the execution phase is completed, the cell has died.
GO:0005739 ; mitochondrion ; A semiautonomous, self replicating organelle that occurs in varying numbers, shapes, and sizes in the cytoplasm of virtually all eukaryotic cells. It is notably the site of tissue respiration.
GO:0006974 ; cellular response to DNA damage stimulus ; Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus indicating damage to its DNA from environmental insults or errors during metabolism.
GO:0006473 ; protein acetylation ; The addition of an acetyl group to a protein amino acid. An acetyl group is CH3CO-, derived from acetic [ethanoic] acid.
GO:0051301 ; cell division ; The process resulting in the physical partitioning and separation of a cell into daughter cells.
GO:0000794 ; condensed nuclear chromosome ; A highly compacted molecule of DNA and associated proteins resulting in a cytologically distinct structure that remains in the nucleus.
GO:0034990 ; nuclear mitotic cohesin complex ; A cohesin complex that mediates sister chromatid cohesion in the nucleus during mitosis; has a subunit composition distinct from that of the meiotic cohesin complex.
GO:0007076 ; mitotic chromosome condensation ; The cell cycle process in which chromatin structure is compacted prior to and during mitosis in eukaryotic cells.
GO:0007059 ; chromosome segregation ; The process in which genetic material, in the form of chromosomes, is organized into specific structures and then physically separated and apportioned to two or more sets. In eukaryotes, chromosome segregation begins with the condensation of chromosomes, includes chromosome separation, and ends when chromosomes have completed movement to the spindle poles.
GO:0000228 ; nuclear chromosome ; A chromosome found in the nucleus of a eukaryotic cell.
GO:0003682 ; chromatin binding ; Interacting selectively and non-covalently

with chromatin, the network of fibers of DNA, protein, and sometimes RNA, that make up the chromosomes of the eukaryotic nucleus during interphase.
GO:0005694 ; chromosome ; A structure composed of a very long molecule of DNA and associated proteins (e.g. histones) that carries hereditary information.
GO:0000775 ; chromosome, centromeric region ; The region of a chromosome that includes the centromeric DNA and associated proteins. In monocentric chromosomes, this region corresponds to a single area of the chromosome, whereas in holocentric chromosomes, it is evenly distributed along the chromosome.
GO:0007049 ; cell cycle ; The progression of biochemical and morphological phases and events that occur in a cell during successive cell replication or nuclear replication events. Canonically, the cell cycle comprises the replication and segregation of genetic material followed by the division of the cell, but in endocycles or syncytial cells nuclear replication or nuclear division may not be followed by cell division.
GO:0006302 ; double-strand break repair ; The repair of double-strand breaks in DNA via homologous and nonhomologous mechanisms to reform a continuous DNA helix.
-------------------------------- RAD27 ----------------------
GO:0017108 ; 5'-flap endonuclease activity ; Catalysis of the cleavage of a 5' flap structure in DNA, but not other DNA structures; processes the 5' ends of Okazaki fragments in lagging strand DNA synthesis.
GO:0090305 ; nucleic acid phosphodiester bond hydrolysis ; The nucleic acid metabolic process in which the phosphodiester bonds between nucleotides are cleaved by hydrolysis.
GO:0003677 ; DNA binding ; Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).
GO:0008409 ; 5'-3' exonuclease activity ; Catalysis of the hydrolysis of ester linkages within nucleic acids by removing nucleotide residues from the 5' end.
GO:0005730 ; nucleolus ; A small, dense body one or more of which are present in the nucleus of eukaryotic cells. It is rich in RNA and protein, is not bounded by a limiting membrane, and is not seen during mitosis. Its prime function is the transcription of the nucleolar DNA into 45S ribosomal-precursor RNA, the processing of this RNA into 5.8S, 18S, and 28S components of ribosomal RNA, and the association of these components with 5S RNA and proteins synthesized outside the nucleolus. This association results in the formation of ribonucleoprotein precursors; these pass into the cytoplasm and mature into the 40S and 60S subunits of the ribosome.
GO:0004527 ; exonuclease activity ; Catalysis of the hydrolysis of ester linkages within nucleic acids by removing nucleotide residues from the 3' or 5' end.
GO:0000287 ; magnesium ion binding ; Interacting selectively and non-covalently with magnesium (Mg) ions.
GO:0005739 ; mitochondrion ; A semiautonomous, self replicating organelle that occurs in varying numbers, shapes, and sizes in the cytoplasm of virtually all eukaryotic cells. It is notably the site of tissue respiration.
GO:0005654 ; nucleoplasm ; That part of the nuclear content other than the chromosomes or the nucleolus.
GO:0006260 ; DNA replication ; The cellular metabolic process in which a cell duplicates one or more molecules of DNA. DNA replication begins when specific sequences, known as origins of replication, are recognized and bound by initiation proteins, and ends when the original DNA molecule has been completely duplicated and the copies topologically separated. The unit of replication usually corresponds to the genome of the cell, an organelle, or a virus. The template for replication can either be an existing DNA molecule or RNA.
GO:0046872 ; metal ion binding ; Interacting selectively and non-covalently

with any metal ion.
GO:0016787 ; hydrolase activity ; Catalysis of the hydrolysis of various bonds, e.g. C-O, C-N, C-C, phosphoric anhydride bonds, etc. Hydrolase is the systematic name for any enzyme of EC class 3.
GO:0006284 ; base-excision repair ; In base excision repair, an altered base is removed by a DNA glycosylase enzyme, followed by excision of the resulting sugar phosphate. The small gap left in the DNA helix is filled in by the sequential action of DNA polymerase and DNA ligase.
GO:0006281 ; DNA repair ; The process of restoring DNA after damage. Genomes are subject to damage by chemical and physical agents in the environment (e.g. UV and ionizing radiations, chemical mutagens, fungal and bacterial toxins, etc.) and by free radicals or alkylating agents endogenously generated in metabolism. DNA is also damaged because of errors during its replication. A variety of different DNA repair pathways have been reported that include direct reversal, base excision repair, nucleotide excision repair, photoreactivation, bypass, double-strand break repair pathway, and mismatch repair pathway.
GO:0016788 ; hydrolase activity, acting on ester bonds ; Catalysis of the hydrolysis of any ester bond.
GO:0006974 ; cellular response to DNA damage stimulus ; Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus indicating damage to its DNA from environmental insults or errors during metabolism.
GO:0003824 ; catalytic activity ; Catalysis of a biochemical reaction at physiological temperatures. In biologically catalyzed reactions, the reactants are known as substrates, and the catalysts are naturally occurring macromolecular substances known as enzymes. Enzymes possess specific binding sites for substrates, and are usually composed wholly or largely of protein, but RNA that has catalytic activity (ribozyme) is often also regarded as enzymatic.
GO:0004518 ; nuclease activity ; Catalysis of the hydrolysis of ester linkages within nucleic acids.
GO:0004519 ; endonuclease activity ; Catalysis of the hydrolysis of ester linkages within nucleic acids by creating internal breaks.
GO:0000737 ; DNA catabolic process, endonucleolytic ; The chemical reactions and pathways resulting in the breakdown of DNA, involving the hydrolysis of internal 3',5'-phosphodiester bonds in one or two strands of deoxyribonucleotides.
GO:0043137 ; DNA replication, removal of RNA primer ; Removal of the Okazaki RNA primer from the lagging strand of replicating DNA, by a combination of the actions of DNA polymerase, DNA helicase and an endonuclease.
GO:0005634 ; nucleus ; A membrane-bounded organelle of eukaryotic cells in which chromosomes are housed and replicated. In most cells, the nucleus contains all of the cell's chromosomes except the organellar chromosomes, and is the site of RNA synthesis and processing. In some species, or in specialized cell types, RNA metabolism or DNA replication may be absent.
------------------------------ AXL2 ----------------------
GO:0005940 ; septin ring ; A tight ring-shaped structure that forms in the division plane at the site of cytokinesis; composed of members of the conserved family of filament-forming proteins called septins as well as septin-associated proteins. This type of septin structure is observed at the bud neck of budding fungal cells, at the site of cell division in animal cells, at the junction between the mother cell and a pseudohyphal projection, and also within hyphae of filamentous fungi at sites where a septum will form.
GO:0005515 ; protein binding ; Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).
GO:0016021 ; integral component of membrane ; The component of a membrane

consisting of gene products and protein complexes that have some part that penetrates at least one leaflet of the membrane bilayer. This component includes gene products that are buried in the bilayer with no exposure outside the bilayer.
GO:0016020 ; membrane ; Double layer of lipid molecules that encloses all cells, and, in eukaryotes, many organelles; may be a single or double lipid bilayer; also includes associated proteins.
GO:0003674 ; molecular_function ; Elemental activities, such as catalysis or binding, describing the actions of a gene product at the molecular level. A given gene product may exhibit one or more molecular functions.
GO:0005509 ; calcium ion binding ; Interacting selectively and non-covalently with calcium ions (Ca2+).
GO:0005887 ; integral component of plasma membrane ; The component of the plasma membrane consisting of gene products and protein complexes that have some part that penetrates at least one leaflet of the membrane bilayer. This component includes gene products that are buried in the bilayer with no exposure outside the bilayer.
GO:0005886 ; plasma membrane ; The membrane surrounding a cell that separates the cell from its external environment. It consists of a phospholipid bilayer and associated proteins.
GO:0005935 ; cellular bud neck ; The constriction between the mother cell and daughter cell (bud) in an organism that reproduces by budding.
GO:0007120 ; axial cellular bud site selection ; The process of defining the next site of bud emergence adjacent to the last site of bud emergence on a budding cell.
-------------------------------- POL30 ----------------------
GO:0030466 ; chromatin silencing at silent mating-type cassette ; Repression of transcription at silent mating-type loci by alteration of the structure of chromatin.
GO:0034087 ; establishment of mitotic sister chromatid cohesion ; The process in which the sister chromatids of a replicated chromosome become joined along the entire length of the chromosome during S phase during a mitotic cell cycle.
GO:0035753 ; maintenance of DNA trinucleotide repeats ; Any process involved in sustaining the fidelity and copy number of DNA trinucleotide repeats. DNA trinucleotide repeats are naturally occurring runs of three base-pairs.
GO:0006298 ; mismatch repair ; A system for the correction of errors in which an incorrect base, which cannot form hydrogen bonds with the corresponding base in the parent strand, is incorporated into the daughter strand. The mismatch repair system promotes genomic fidelity by repairing base-base mismatches, insertion-deletion loops and heterologies generated during DNA replication and recombination.
GO:0003677 ; DNA binding ; Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid).
GO:0007064 ; mitotic sister chromatid cohesion ; The cell cycle process in which the sister chromatids of a replicated chromosome are joined along the entire length of the chromosome, from their formation in S phase through metaphase during a mitotic cell cycle. This cohesion cycle is critical for high fidelity chromosome transmission.
GO:0006348 ; chromatin silencing at telomere ; Repression of transcription of telomeric DNA by altering the structure of chromatin.
GO:0006301 ; postreplication repair ; The conversion of DNA-damage induced single-stranded gaps into large molecular weight DNA after replication. Includes pathways that remove replication-blocking lesions in conjunction with DNA replication.
GO:0005634 ; nucleus ; A membrane-bounded organelle of eukaryotic cells in which chromosomes are housed and replicated. In most cells, the nucleus contains all of the cell's chromosomes except the organellar chromosomes, and is the site of RNA synthesis and processing. In some species, or in

specialized cell types, RNA metabolism or DNA replication may be absent.
GO:0005657 ; replication fork ; The Y-shaped region of a replicating DNA molecule, resulting from the separation of the DNA strands and in which the synthesis of new strands takes place. Also includes associated protein complexes.
GO:0006260 ; DNA replication ; The cellular metabolic process in which a cell duplicates one or more molecules of DNA. DNA replication begins when specific sequences, known as origins of replication, are recognized and bound by initiation proteins, and ends when the original DNA molecule has been completely duplicated and the copies topologically separated. The unit of replication usually corresponds to the genome of the cell, an organelle, or a virus. The template for replication can either be an existing DNA molecule or RNA.
GO:0042802 ; identical protein binding ; Interacting selectively and non-covalently with an identical protein or proteins.
GO:0030337 ; DNA polymerase processivity factor activity ; An enzyme regulator activity that increases the processivity of polymerization by DNA polymerase, by allowing the polymerase to move rapidly along DNA while remaining topologically bound to it.
GO:0000278 ; mitotic cell cycle ; Progression through the phases of the mitotic cell cycle, the most common eukaryotic cell cycle, which canonically comprises four successive phases called G1, S, G2, and M and includes replication of the genome and the subsequent segregation of chromosomes into daughter cells. In some variant cell cycles nuclear replication or nuclear division may not be followed by cell division, or G1 and G2 phases may be absent.
GO:0006281 ; DNA repair ; The process of restoring DNA after damage. Genomes are subject to damage by chemical and physical agents in the environment (e.g. UV and ionizing radiations, chemical mutagens, fungal and bacterial toxins, etc.) and by free radicals or alkylating agents endogenously generated in metabolism. DNA is also damaged because of errors during its replication. A variety of different DNA repair pathways have been reported that include direct reversal, base excision repair, nucleotide excision repair, photoreactivation, bypass, double-strand break repair pathway, and mismatch repair pathway.
GO:0006289 ; nucleotide-excision repair ; A DNA repair process in which a small region of the strand surrounding the damage is removed from the DNA helix as an oligonucleotide. The small gap left in the DNA helix is filled in by the sequential action of DNA polymerase and DNA ligase. Nucleotide excision repair recognizes a wide range of substrates, including damage caused by UV irradiation (pyrimidine dimers and 6-4 photoproducts) and chemicals (intrastrand cross-links and bulky adducts).
GO:0006974 ; cellular response to DNA damage stimulus ; Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus indicating damage to its DNA from environmental insults or errors during metabolism.
GO:0043626 ; PCNA complex ; A protein complex composed of three identical PCNA monomers, each comprising two similar domains, which are joined in a head-to-tail arrangement to form a homotrimer. Forms a ring-like structure in solution, with a central hole sufficiently large to accommodate the double helix of DNA. Originally characterized as a DNA sliding clamp for replicative DNA polymerases and as an essential component of the replisome, and has also been shown to be involved in other processes including Okazaki fragment processing, DNA repair, translesion DNA synthesis, DNA methylation, chromatin remodeling and cell cycle regulation.
GO:0006273 ; lagging strand elongation ; The synthesis of DNA from a template strand in a net 3' to 5' direction. Lagging strand DNA elongation proceeds by discontinuous synthesis of short stretches of DNA, known as Okazaki fragments, from RNA primers; these fragments are then joined by DNA ligase. Although each segment of nascent DNA is synthesized in the 5' to 3'

direction, the overall direction of lagging strand synthesis is 3' to 5', mirroring the progress of the replication fork.
GO:0006272 ; leading strand elongation ; The synthesis of DNA from a template strand in the 5' to 3' direction; leading strand elongation is continuous as it proceeds in the same direction as the replication fork.
GO:0000710 ; meiotic mismatch repair ; A system for the identification and correction of base-base mismatches, small insertion-deletion loops, and regions of heterology that are present in duplex DNA formed with strands from two recombining molecules. Correction of the mismatch can result in non-Mendelian segregation of alleles following meiosis.
GO:0006275 ; regulation of DNA replication ; Any process that modulates the frequency, rate or extent of DNA replication.
GO:0070987 ; error-free translesion synthesis ; The conversion of DNA-damage induced single-stranded gaps into large molecular weight DNA after replication by using a specialized DNA polymerase or replication complex to insert a defined nucleotide across the lesion. This process does not remove the replication-blocking lesions but does not causes an increase in the endogenous mutation level. For S. cerevisiae, RAD30 encodes DNA polymerase eta, which incorporates two adenines. When incorporated across a thymine-thymine dimer, it does not increase the endogenous mutation level.
GO:0005515 ; protein binding ; Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).
GO:1902394 ; positive regulation of exodeoxyribonuclease activity ; Any process that activates or increases the frequency, rate or extent of exodeoxyribonuclease activity.