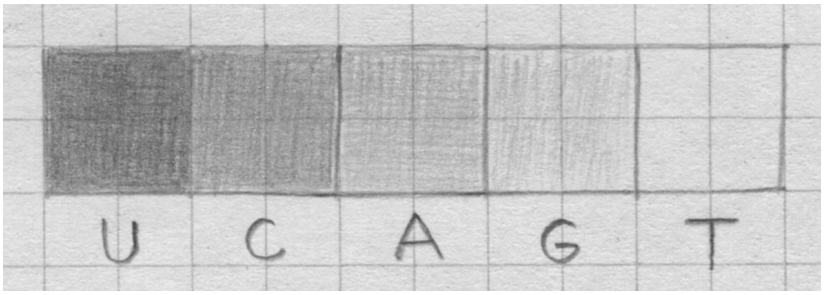
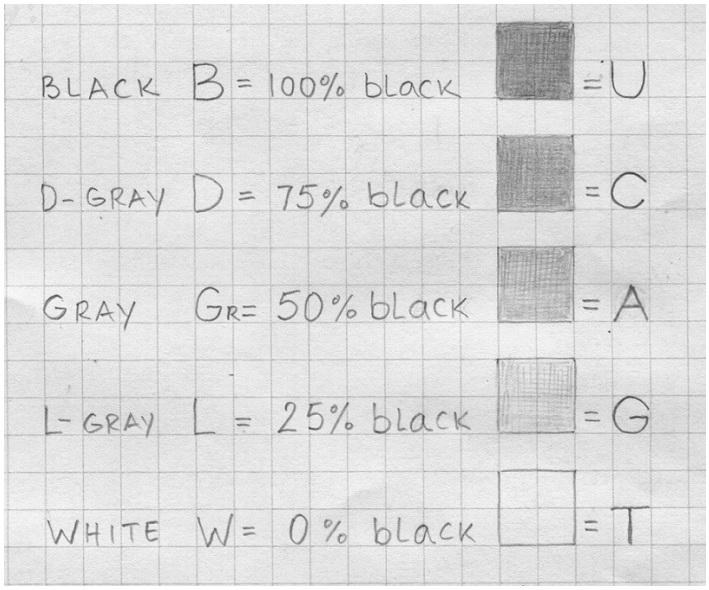
Numeric Representation of DNA and RNA

Widely accepted representation of RNA/DNA bases are five initial letters of their Latin names Thymine=T, Guanine=G, Adenine=A, Cytosine=C and Uracil=U. Based on these, in essence arbitrary, letters this representation is the least revealing and functional. From these letters we could only guess the name of the base, but wont be able to learn anything about the relationships between the bases.

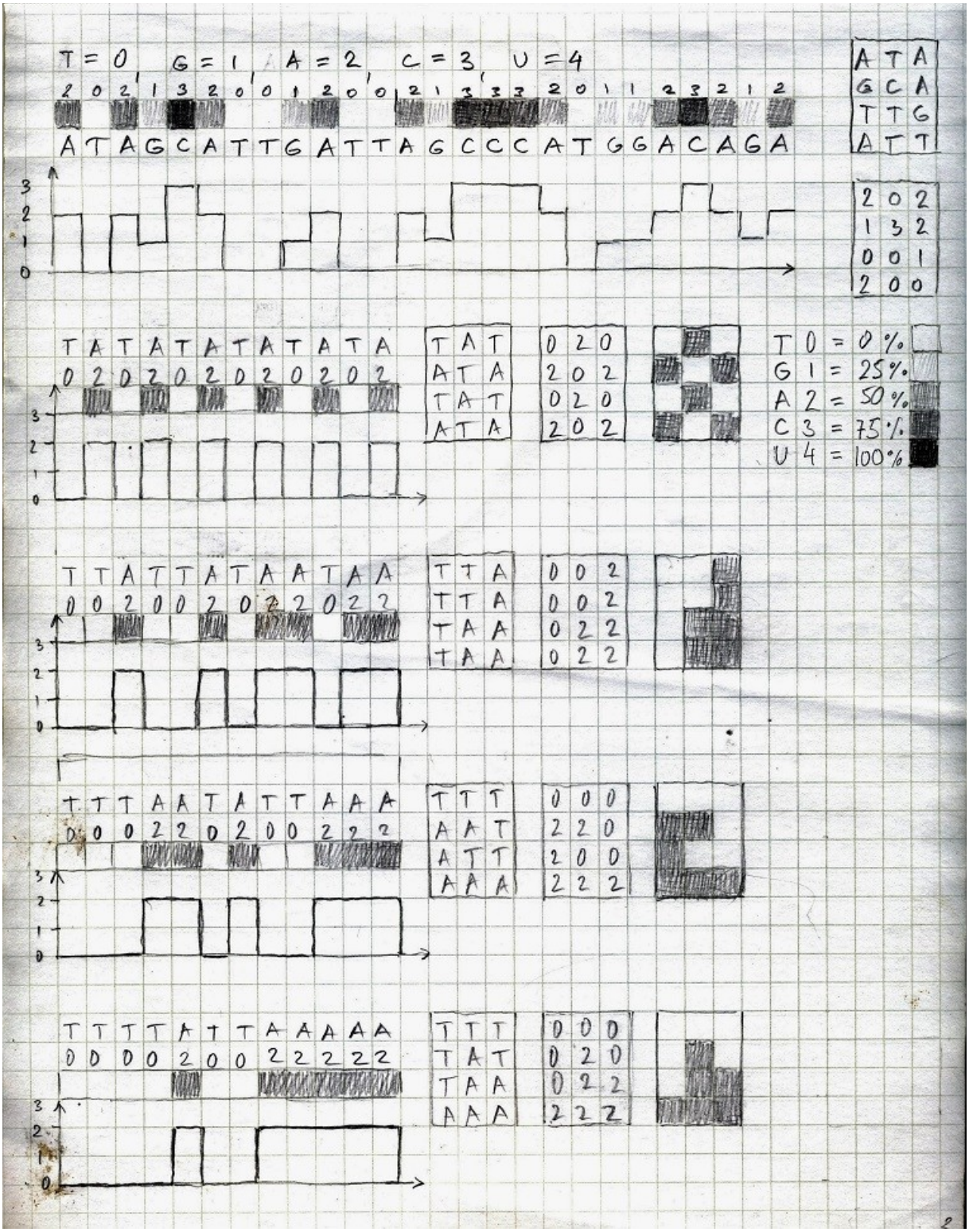


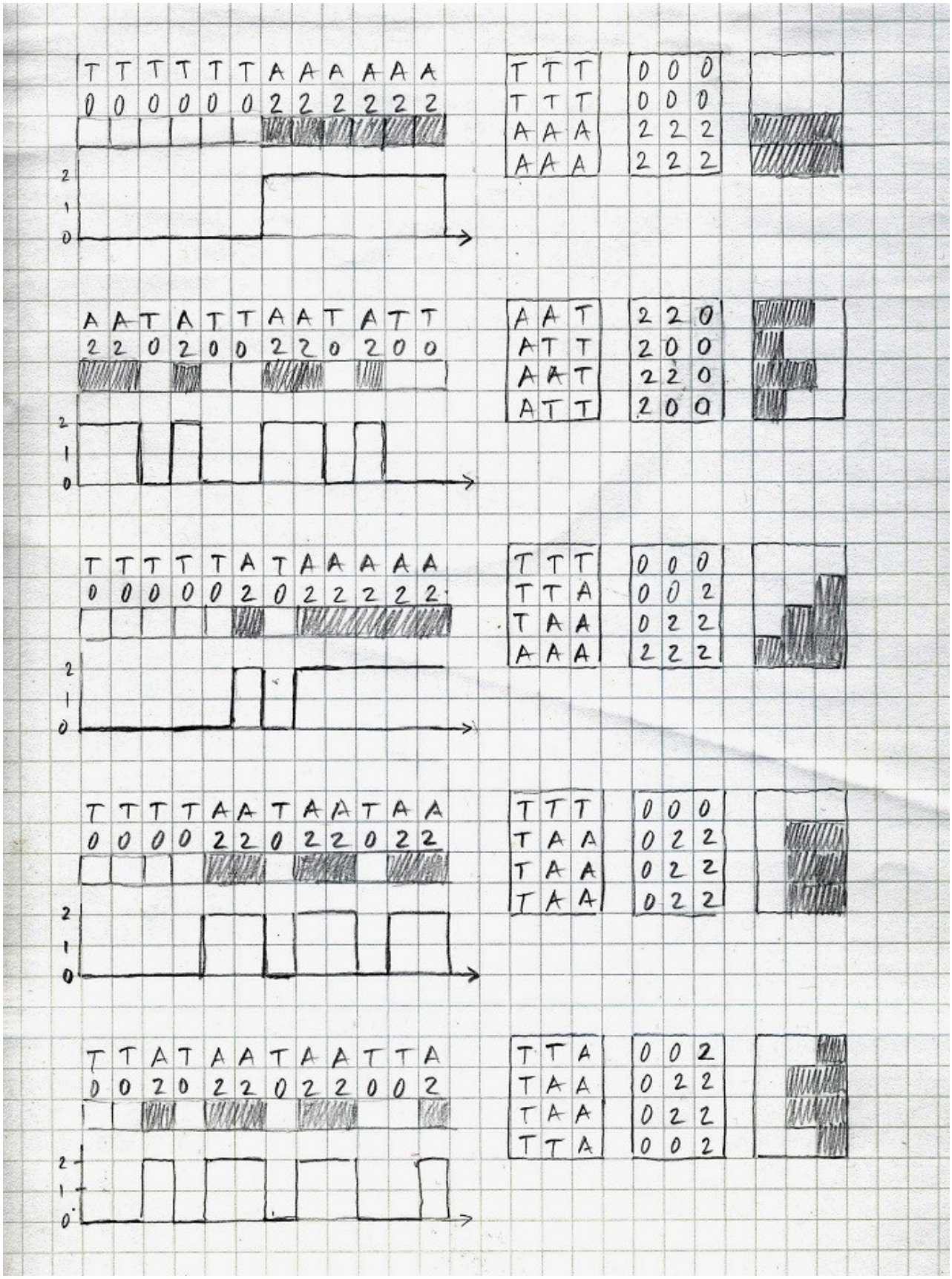
Another representation based on five values from the gray scale where T=White(0% black), G=Light(25% black), A=Gray(50% black), C= Dark (75% black) and U=Black(100% black) much better reflects some properties of RNA and DNA. Instead of listing all three base-pairs T-A, G-C and A-U, we could look for all the pairs with 50% value difference and we will get only these three pairs. In addition, we could notice that the base-pair relationship is not neutral, but in case G-C it is addition(G+50%=C) while C-G is a subtraction(C-50%=G). It might be interesting to notice that in all six base-pair transactions the exchange value is the same 50% black that is the symbolic value for Adenine. It seems as if Adenine represents an exchange currency in these transactions.



Based on this visual representation it is possible to define a corresponding numeric representation of the bases either expressed as percentage values as mentioned above: T=0%, G=25%, A=50%, C=75% , U=100% or as whole numbers T=0, G=1, A=2, C=3, U=4(DNA 0,1,2, 3 and RNA 1,2,3, 4). Here the exchange value for the base-pairs is 2 which is also the number that represents Adenine. Thus, instead of a DNA sequence expressed through alphabet letters ATAGCATTGATTAGCCCATGGACAGA we could have its numeric version as 202132001200213332011232l2. This numeric representation could be presented in 2D as a Cartesian diagram where on the **x** are positions of the bases while on the **y** are their values. Below are cases of the sequences with high and low entropy as well as a few stretches of human chromosome Chr22 represented in the numeric form.

Gregor Mobius New York, June 15, 2019

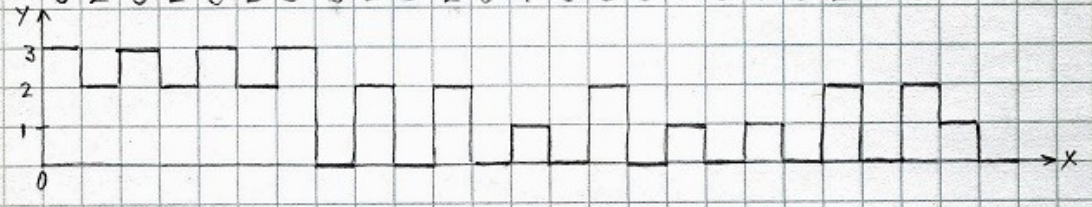

T = 0 G = 1 A = 2 C = 3 U = 4
2 0 2 1 3 2 0 0 1 2 0 0 2 1 3 3 3 2 0 1 1 2 3 2 1 2
A T A G C A T T G A T T A G C C C A T G G A C A G A
A T A
G C A
T T G
A T T
2 0 2
1 3 2
0 0 1
2 0 0
T A T A T A T A T A T A T A
0 2 0 2 0 2 0 2 0 2 0 2 0 2
T A T
A T A
T A T
A T A
0 2 0
2 0 2
0 2 0
2 0 2
T 0 = 0 %
G 1 = 25 %
A 2 = 50 %
C 3 = 75 %
U 4 = 100 %
T T A T T A T A A T A A
0 0 2 0 0 2 0 2 2 0 2 2
T T A
T T A
T A A
T A A
0 0 2
0 0 2
0 2 2
0 2 2
T T T A A T A T T A A A
0 0 0 2 2 0 2 0 0 2 2 2
T T T
A A T
A T T
A A A
0 0 0
2 2 0
2 0 0
2 2 2
T T T T A T T A A A A A
0 0 0 0 2 0 0 2 2 2 2 2
T T T
T A T
T A A
A A A
0 0 0
0 2 0
0 2 2
2 2 2


T T T T T T A A A A A A
0 0 0 0 0 0 2 2 2 2 2 2
2
1
0
T T T
T T T
A A A
A A A
0 0 0
0 0 0
2 2 2
2 2 2
A A T A T T A A T A T T
2 2 0 2 0 0 2 2 0 2 0 0
2
1
0
A A T
A T T
A A T
A T T
2 2 0
2 0 0
2 2 0
2 0 0
T T T T T A T A A A A A
0 0 0 0 0 2 0 2 2 2 2 2
2
1
0
T T T
T T A
T A A
A A A
0 0 0
0 0 2
0 2 2
2 2 2
T T T T A A T A A T A A
0 0 0 0 2 2 0 2 2 0 2 2
2
1
0
T T T
T A A
T A A
T A A
0 0 0
0 2 2
0 2 2
0 2 2
T T A T A A T A A T T A
0 0 2 0 2 2 0 2 2 0 0 2
2
1
0
T T A
T A A
T A A
T T A
0 0 2
0 2 2
0 2 2
0 0 2

HUMAN CHROMOSOME CHR 22

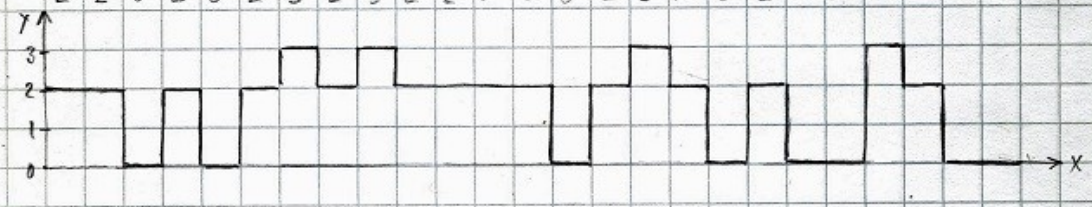
P. 15221470

	C	A	C	A	C	A	C	T	A	T	A	T	G	T	A	T	G	T	G	T	A	T	A	G	T
%	75	50	75	50	75	50	75	0	50	0	50	0	25	0	50	0	25	0	25	0	50	0	50	25	0
	3	2	3	2	3	2	3	0	2	0	2	0	1	0	2	0	1	0	1	0	2	0	2	1	0



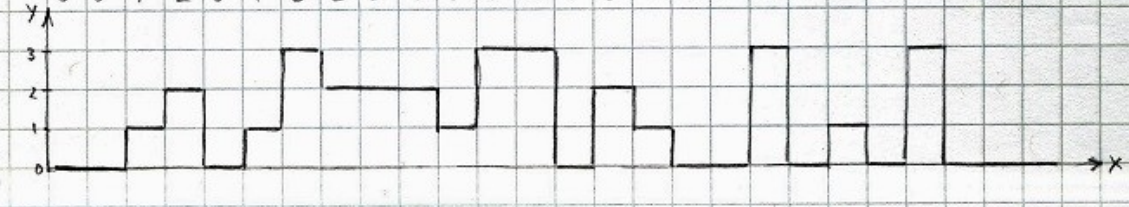
P. 15244582

	A	A	T	A	T	A	C	A	C	A	A	A	A	T	A	C	A	T	A	T	T	C	A	T	T
%	50	50	0	50	0	50	75	50	75	50	50	50	50	0	50	75	50	0	50	0	0	75	50	0	0
	2	2	0	2	0	2	3	2	3	2	2	2	2	0	2	3	2	0	2	0	0	3	2	0	0



P. 15251768

	T	T	G	A	T	G	C	A	A	A	G	C	C	T	A	G	T	T	C	T	G	T	C	T	T	T
%	0	0	25	50	0	25	75	50	50	50	25	75	75	0	50	25	0	0	75	0	25	0	75	0	0	0
	0	0	1	2	0	1	3	2	2	2	1	3	3	1	2	1	0	0	3	0	1	0	3	0	0	0



P. 15252038

	G	T	T	C	T	A	A	G	T	T	T	G	T	T	C	C	A	G	G	G	T	C	T	C	A
	25	0	0	75	0	50	50	25	0	0	0	25	0	0	75	75	50	25	25	25	0	75	0	75	50
	1	0	0	3	0	2	2	1	0	0	0	1	0	0	3	3	2	1	1	1	0	3	0	3	2