1 The length scales of DNA

The genome of an organism is the sequence of DNA base-pairs (A, T, C and G) that this organism contains. The DNA in human cells is organised in macromolecules called chromosomes. The human genome contains 22 chromosomes plus 2 sexual chromosomes, called X and Y. Its length is nowdays estimated at $3,2 \cdot 10^9$ base-pairs.
The large majority of human cells are diploids. It means that they contain two copies of each non-sexual chromosome plus two sexual chromosomes, that is, $2 \cdot 22 + 2 = 46$ chromosomes. Thus, a diploid cell contains two times the genome, that is, approximately $6,4 \cdot 10^9$ base pairs of DNA.

1. If T is the total number of base-pairs in DNA and h is the base-pair length, then the total length L of DNA in a cell is given by

$$\begin{aligned} L &= T \cdot h \\ &= 6 \cdot 10^9 \cdot 3,4 \cdot 10^{-10}\,\mathrm{m} = 6 \cdot 3,4 \cdot 10^{-1}\,\mathrm{m} \\ &= 2.04\,\mathrm{m} \end{aligned} \tag{1.1}$$

Thus the total length of DNA in a cell is approximately 2 meters, or $\cdot 10^5$ times the diameter of the cell. The DNA must therefore be highly folded!

2.

$$r = \frac{(\text{DNA diameter})}{2} = \frac{20}{2}\text{Å} = 10^{-9}\text{m} \tag{1.2}$$

$$R = \frac{(\text{cell diameter})}{2} = \frac{10^5}{2}\text{Å} = 5 \cdot 10^{-6}\text{m} \tag{1.3}$$

The volume of DNA is given by

$$\begin{aligned} v &= L \cdot \pi \cdot r^2 \\ &= 2,04 \cdot \pi \cdot 10^{-18}\,\mathrm{m}^3 \\ &= 6,41 \cdot 10^{-18}\mathrm{m}^3 \end{aligned} \tag{1.4}$$

The volume of the cell is given by

$$\begin{aligned} V &= \frac{4}{3} \cdot \pi \cdot R^3 \\ &= \frac{4}{3} \cdot \pi \cdot 1,25 \cdot 10^{-16}\,\mathrm{m}^3 \\ &= 5,23 \cdot 10^{-16}\,\mathrm{m}^3 \end{aligned} \tag{1.5}$$

so that

$$\frac{V}{v} = 81,68 \tag{1.6}$$

The DNA volume is thus 81 times smaller than the cell volume! That is it fills only about 1.22% of the available volume. (In point of fact in eukaryotes the DNA is stored inside the nucleus of the cell so that the available volume is substantially smaller than that of the cell itself. But nevertheless there is plenty of volume available to store the DNA—the problem is how to coil such a long object in an organized fashion to be packed.)

For those who are interested in more details about the human genome: The human genome sequence as been published in Science [1] in 2001. 2.91-billion base pair (bp) of the euchromatic portion of the human genome were sequenced, including the sequence of 22 chromosomes and the sexual chromosomes X (155 millions bp) and Y (58 millions bp).
A typical human cell is dioploidic which means it has two copies of each non-sexual chromosome, so it has for a male $(2.91-\mathrm{x}-\mathrm{y})\cdot 2+\mathrm{x}+\mathrm{y}=2\cdot 2.91-\mathrm{x}-\mathrm{y}=(2\cdot 2.9$ billions-155 millions -58millions$=5587$ millions base pairs of DNA, and for a female $(2.91-\mathrm{x}-\mathrm{y})\cdot 2+2\mathrm{x}=2\cdot 2.91-2\mathrm{y}=5684$ millions bp.
The cell also contains 10% of another type of DNA called heterochromatin[2] that was not sequenced in the Science publication, so the total number of base pairs has to be increased by ten percent: $5.6\cdot 1.1$ billions ≈ 6 billions bp.
Before mitosis (cell division) all the DNA is replicated, which doubles the total amount of DNA in the cell (around 12 billion bp).
A base pair junction is 3.4 Angstrom $(0.34\cdot 10-9\mathrm{m})$ long so the total DNA length varies between $6\cdot 10^9\cdot 0.34\cdot 10^{-9}$ and $12\cdot 10^9\cdot 0.34\cdot 10^{-9}$ meters, that is between 2 to 4 meters depending on the phase in the cell cycle.
The cell has also mitochondrial DNA in the cytoplasm but it is negligible (only a few tens of thousands bp).

2 Gaussian integrals

The matrix K is symetric, positive-definite, so an orthonormal basis exists in which K is diagonal.

1. Change to polar coordinates

$$\begin{aligned}\left(\int_{-\infty}^{\infty}e^{-\sigma^2}d\sigma\right)^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-(\sigma^2+\tau^2)}d\sigma d\tau=\int_0^{\infty}\int_0^{2\pi}e^{-r^2}rd\phi dr\\ &= 2\pi\int_0^{\infty}e^{-r^2}rdr=2\pi\lim_{R\to\infty}-\frac{1}{2}e^{-r^2}\Big|_0^R=\pi.\end{aligned}$$

2. Using change of variables ($x=Py$ such that $P^TKP=diag(\lambda_1,...,\lambda_n)$, $\det(P)=1$) we get

$$\begin{aligned}Z^{(n)}:=\int_{R^n}e^{-\beta xKx}dx &= \int_{R^n}e^{-\sum_{i=1}^n\beta\lambda_i y_i^2}dy=\prod_{i=1}^n\int_{-\infty}^{\infty}e^{-\beta\lambda_i y_i^2}dy_i\\ &= \prod_{i=1}^n\sqrt{\frac{1}{\beta\lambda_i}}\int_{-\infty}^{\infty}e^{-y_i^2}dy_i=\prod_{i=1}^n\sqrt{\frac{\pi}{\beta\lambda_i}}\\ &= \left(\frac{\pi}{\beta}\right)^{\frac{n}{2}}\sqrt{\det[K^{-1}]}.\end{aligned}$$

3. The first integral is zero because $x_ie^{-\beta xKx}$ is an odd function. We will reduce this to the one-dimentional case. Let $1\le i\le n$. As in 2 change basis and write

$$x_i=\sum_{k=1}^n p_{ik}y_k.$$

Then

$$\int_{R^n}x_ie^{-\beta x\cdot Kx}\,dx=\int_{R^n}\sum_{k=1}^n p_{ik}y_ke^{-\sum_{i=1}^n\beta\lambda_i y_i^2}\,dy$$

[1] The Sequence of the Human Genome Science 16 February 2001 Vol. 291. no. 5507, pp. 1304 - 1351
[2] Alberts et al, Molecular Biology of the Cell third edition, Garland Publishing, 1994, pp 353

$$
\begin{aligned}
&= \sum_{k=1}^{n} p_{ik} \int_{R^n} y_k e^{-\sum_{i=1}^{n} \beta\lambda_i y_i^2}\, dy \\
&= \sum_{k=1}^{n} p_{ik} \left(\prod_{r=1}^{k-1} \int_{-\infty}^{\infty} e^{-\beta\lambda_r y_r^2}\, dy_r \right) \int_{-\infty}^{\infty} y_k e^{-\beta\lambda_k y_k^2}\, dy_k \left(\prod_{t=k+1}^{n} \int_{-\infty}^{\infty} e^{-\beta\lambda_t y_t^2}\, dy_t \right) \\
&= 0,
\end{aligned}
$$

because $y_k e^{-\beta\lambda_k y_k^2}$ is an odd function. For the second integral we use a change of variables $u = x - \hat{x}$. Then

$$
\int_{R^n} x_i e^{-\beta(x-\hat{x})\cdot K(x-\hat{x})}\, dx = \int_{R^n} (u_i + \hat{x}_i) e^{-\beta u \cdot Ku}\, dx = Z^{(n)} \hat{x}_i.
$$

4. For the one-dimensional case we get

$$
\int_{-\infty}^{\infty} y_i^2 e^{-\beta\lambda_i y_i^2} dy_i = \frac{1}{2\beta\lambda_i} \sqrt{\frac{\pi}{\beta\lambda_i}}, \tag{2.1}
$$

using a similar technique as in 1.

For more-dimensional case we change basis (similar to 2) and write

$$
x_i = \sum_{k=1}^{n} p_{ik} y_k, \quad x_j = \sum_{l=1}^{n} p_{jl} y_l.
$$

Again we insert this into the integral

$$
\begin{aligned}
\int_{R^n} x_i x_j e^{-\beta x \cdot Kx} dx &= \sum_{k,l=1}^{n} p_{ik} p_{jl} \int_{R^n} y_k y_l e^{-\sum_{k=1}^{n} \beta\lambda_i y_i^2} dy \\
&= \sum_{k=1}^{n} p_{ik} p_{jk} \int_{R^n} y_k^2 e^{-\sum_{i=1}^{n} \beta\lambda_i y_i^2} dy \qquad (2.2) \\
&+ \sum_{k\neq l} p_{ik} p_{jl} \int_{R^n} y_k y_l e^{-\sum_{i=1}^{n} \beta\lambda_i y_i^2} dy.
\end{aligned}
$$

By the same argument as in 3, namely as the function $y_k e^{-\beta\lambda_k y_k^2}$ is odd, the second sum in (2.2) is zero. To calculate the first sum we use (2.1) and 1:

$$
\begin{aligned}
\int_{R^n} x_i x_j e^{-\beta x \cdot Kx} dx &= \sum_{k=1}^{n} p_{ik} p_{jk} \frac{1}{2\beta\lambda_k} \left(\frac{\pi}{\beta} \right)^{\frac{n}{2}} \sqrt{\det[K^{-1}]} \\
&= \frac{Z^{(n)}}{2\beta} \sum_{k=1}^{n} p_{ik} \frac{1}{\lambda_k} p_{jk}
\end{aligned}
$$

Finally we have just to remark that $K^{-1} = P\, diag(\lambda_1^{-1}, \ldots, \lambda_n^{-1})\, P^T$ and that $[K^{-1}]_{ij} = \sum_{k=1}^{n} p_{ik} \frac{1}{\lambda_k} p_{jk}$.

5. We first recall the definition of the Forbenius inner product: $A : B := \sum_{i,j=1}^{n} a_{ij}b_{ij} = \mathrm{tr}(B^T A) = \mathrm{tr}(A^T B)$, and the invariance of the trace operator under cyclic permutation: $\mathrm{tr}(ABC) = \mathrm{tr}(CAB) = \mathrm{tr}(BCA)$, for three matrices A, B, C with compatible dimensions. We have then that:

$$\begin{aligned}
\left\langle (x-\widehat{x})^T K(x-\widehat{x}) \right\rangle_p &= \left\langle \mathrm{tr}\left((x-\widehat{x})^T K(x-\widehat{x})\right) \right\rangle_p && (2.3)\\
&= \left\langle K : (x-\widehat{x})(x-\widehat{x})^T \right\rangle_p && (2.4)\\
&= K : \left\langle (x-\widehat{x}_(x-\widehat{x})^T \right\rangle_p && (2.5)\\
&= K : K^{-1} = \mathrm{tr}\left(KK^{-1}\right) = n. && (2.6)
\end{aligned}$$

Remark

There is an easiest way to compute the n-dimensional Gaussian integrals in (3) and (4) that uses only the knowledge of the uni-dimensional case. We stress that in (3) and in (4) we actually compute the first and the second (centred and non–centred) moments of a multivariate Gaussian distribution.
Let us define the Gaussian measure as

$$d\rho(\mathbf{x}) = \frac{1}{Z^{(n)}} e^{-\beta U(\mathbf{x})} d\mathbf{x},\ \mathbf{x} \in I\!R^n,\ K = K^T, K > 0, \tag{2.7}$$

with $U(\mathbf{x}) = \dfrac{1}{2}\mathbf{x}^T K\mathbf{x}$, $K = K^T$, and $K > 0$. Without lost of generality we can assume $\beta = 1$, which implies that the quadratic energy $U(\mathbf{x})$ is already non-dimensional. The expectation of an observable $f(\mathbf{x})$ with respect to that measure $d\rho(\mathbf{x})$ is defined by

$$\langle f(\mathbf{x}) \rangle := \int_{R^n} f(\mathbf{x}) d\rho = \frac{1}{Z^{(n)}} \int_{R^n} f(\mathbf{x}) e^{-\frac{1}{2}\mathbf{x}\cdot K\mathbf{x}}\, d\mathbf{x}. \tag{2.8}$$

Using the properties of K we can now introduce the same change of variable done in the previous parts, namely, $\mathbf{x} = P\mathbf{y}$, where $P^T K P = diag(\lambda_1, \ldots, \lambda_n) = \Lambda$. Then the measure (2.7) becames

$$d\rho(\mathbf{y}) = \frac{1}{Z^{(n)}} e^{-\frac{1}{2}\mathbf{y}\cdot\Lambda\mathbf{y}} d\mathbf{y}, = \frac{1}{Z^{(n)}} \prod_{i=1}^{n} e^{-\frac{1}{2}\lambda_i y_i^2} dy_i, \tag{2.9}$$

and we have just restrict our self to the scalar case. We stress here that the partition function $Z^{(n)}$ do split into a product of one-dimensional partition functions $Z_i^{(1)} = \sqrt{\frac{2\pi}{\lambda_i}}$. Then just by knowing the following results of the scalar Gaussian integrals

$$\begin{aligned}
\frac{1}{Z^{(1)}} \int_{-\infty}^{+\infty} y e^{-\frac{1}{2}\lambda y^2} dy &= 0,\\
\frac{1}{Z^{(1)}} \int_{-\infty}^{+\infty} y^2 e^{-\frac{1}{2}\lambda y^2} dy &= \frac{1}{\lambda}
\end{aligned}$$

and the basic property of the expectation, we can then compute easily the first two moments of the multivariate Gaussian distribution. We recall that $\langle \mathbf{x} \rangle$ and $\langle \mathbf{x} \otimes \mathbf{x} \rangle$ are, respectively, the first and the second centred moments. Finally we obtain

$$\begin{aligned}
\langle \mathbf{x} \rangle &= \langle P\mathbf{y} \rangle = P \langle \mathbf{y} \rangle = P\mathbf{0} = \mathbf{0},\\
\langle \mathbf{x} \otimes \mathbf{x} \rangle &= \langle P\mathbf{y} \otimes P\mathbf{y} \rangle = P \langle \mathbf{y} \otimes \mathbf{y} \rangle P^T = P\Lambda^{-1}P^T = K^{-1}.
\end{aligned}$$

The latter proceeding works also with shifted quadratic energies ,i.e, energies of the form $U(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^T K(\mathbf{x} - \hat{\mathbf{x}})$ for a given $\hat{\mathbf{x}} \in \mathbb{R}^n$. The only change is that now the expectation $\langle \mathbf{x} \rangle = \hat{\mathbf{x}}$.