(a).

$$F(q) = E_q[\log p(x|z)] - D_{KL}(q(z) \| p(z))$$

$$= E_q\left[\log \frac{p(z|x)p(x)}{p(z)}\right] - E_q[\log q(z) - \log p(z)]$$

$$= E_q[\log p(z|x) + \log p(x) - \cancel{\log p(z)} - \log q(z) + \cancel{\log p(z)}]$$

$$= E_q[\log p(x)] - E_q[\log q(z) - \log p(z|x)]$$

$$= \log p(x) - D_{KL}(q(z) \| p(z|x))$$

(b)

$$D_{KL}(q(z) \,||\, p(z))$$

$$= E_q[\log q(z) - \log p(z)]$$

$$= E_q[\log \prod_{i=1}^{D} q_i(z_i) - \log \prod_{i=1}^{D} p_i(z_i)]$$

$$= E_q[\sum_{i=1}^{D} \log q_i(z_i) - \sum_{i=1}^{D} \log p_i(z_i)]$$

$$= \sum_{i=1}^{D} E_{q_i}[\log q_i(z_i) - \log p_i(z_i)]$$

$$= \sum_{i=1}^{D} D_{KL}(q_i(z_i) \,||\, p_i(z_i))$$

(c).

$$D_{KL}(q_i(z_i) \,||\, p_i(z_i))$$

$$= E_{q_i}[\log q_i(z) - \log p_i(z)]$$

$$= E_{q_i}\left[\log \frac{\frac{1}{\cancel{\sqrt{2\pi}}\sigma_i} \exp\{-\frac{(z_i-\mu_i)^2}{2\sigma_i^2}\}}{\frac{1}{\cancel{\sqrt{2\pi}}} \exp\{-\frac{z_i^2}{2}\}}\right]$$

$$= E_{q_i}\left[\log \frac{1}{\sigma_i} \exp\{-\frac{(z_i-\mu_i)^2}{2\sigma_i^2}\} - \log \exp\{-\frac{z_i^2}{2}\}\right]$$

$$= E_{q_i}\left[-\log \sigma_i - \frac{(z_i-\mu_i)^2}{2\sigma_i^2} + \frac{z_i^2}{2}\right]$$

$$= -\log \sigma_i - \frac{1}{2} E_{q_i}\left(\frac{(z_i-\mu_i)^2}{\sigma_i^2}\right) + \frac{1}{2} E_{q_i}(z_i^2)$$

$$= -\log \sigma_i - \frac{1}{2}(1+0^2) + \frac{1}{2}(\sigma_i^2+\mu_i^2)$$

$$= -\frac{1}{2} - \log \sigma_i + \frac{1}{2}(\sigma_i^2+\mu_i^2)$$

(d).

$$E_\varepsilon \nabla_\theta t_i = E_\varepsilon \begin{pmatrix} \frac{\partial t_i}{\partial \mu_i} \\ \frac{\partial t_i}{\partial \sigma_i} \end{pmatrix}$$

$$E_\varepsilon\left(\frac{\partial t_i}{\partial \mu_i}\right) = E_\varepsilon\left[\frac{\partial\left(\log \frac{q_i(z_i)}{p_i(z_i)}\right)}{\partial \mu_i}\right]$$

$$= E_\varepsilon\left[\frac{\partial\left(-\log \sigma_i - \frac{(z_i-\mu_i)^2}{2\sigma_i^2} + \frac{z_i^2}{2}\right)}{\partial \mu_i}\right] \Leftarrow z_i = \mu_i + \sigma_i \varepsilon_i$$

$$= E_\varepsilon\left[0 + \frac{\partial \frac{(\cancel{\mu_i} + \cancel{\sigma_i}\varepsilon_i - \cancel{\mu_i})^2}{\cancel{2\sigma_i^2}}}{\partial \mu_i} + (\mu_i + \sigma_i \varepsilon_i)\right]$$

$$= E_\varepsilon\left[\frac{1}{2}\frac{\partial \varepsilon_i^2}{\partial \mu_i} + \mu_i + \sigma_i \varepsilon_i\right]$$

$$= 0 + \mu_i + \sigma_i \cancel{E_\varepsilon(\varepsilon_i)}^{\,0}$$

$$= \mu_i$$

$$E_\varepsilon\left(\frac{\partial t_i}{\partial \sigma_i}\right) = E_\varepsilon\left[-\frac{1}{\sigma_i} + \frac{\dfrac{\partial(\cancel{\mu_i} + \cancel{\sigma_i}\varepsilon_i - \cancel{\mu_i})^2}{\partial\cancel{\sigma_i^2}}}{\partial\sigma_i} + \varepsilon_i(\mu_i+\sigma_i\varepsilon_i)\right]$$

$$= E_\varepsilon\left[-\frac{1}{\sigma_i} + \frac{\partial \varepsilon_i^2}{\partial\sigma_i} + \varepsilon_i(\mu_i+\sigma_i\varepsilon_i)\right]$$

$$= -\frac{1}{\sigma_i} + \sigma_i$$

$$\nabla_\theta D_{KL}(q_i(z_i)\,\|\,p_i(z_i)) = E_\varepsilon[\nabla_\theta t_i]$$

$$= \begin{pmatrix} \mu_i \\ \\ \sigma_i - \dfrac{1}{\sigma_i} \end{pmatrix}$$